(2)

OPERATIONS RESEARCH CENTER

A GENERAL MODEL OF PRODUCTION: THEORY AND APPLICATION

by

Steven T. Hackman

ORC 84-2

January 1984

UNIVERSITY OF CALIFORNIA

BERKELEY

A GENERAL MODEL OF PRODUCTION:
THEORY AND APPLICATION

by

Steven T. Hackman[†]

ORC 84-2

January 1984

[†]Operations Research Center, University of California, Berkeley, California.

This research was supported by the Office of Naval Research under Contract N00014-76-C-0134 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC 84-2	2. GOVT ACCESSION NO. AD-A158020	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A GENERAL MODEL OF PRODUCTION: THEORY AND APPLICATION		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Steven T. Hackman		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0134
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 337 015
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, Virginia 22217		12. REPORT DATE January 1984
		13. NUMBER OF PAGES 150
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Production; Production Model; Production Networks; Shephard's Duality; Laws of Diminishing Returns; Aggregation; Multi-Project Resource-Use Planning; Materials Requirement Planning; Critical Path Method; Production Function

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DEDICATION

To

Mom and Dad

and

Audrey, Mike, Amy, and Mindy

Acknowledgements

It is difficult to adequately express my deep felt appreciation to my committee chairman and friend, Professor Robert C. Leachman. I will just say that he has been, in every conceivable way, a superb thesis advisor.

The late Professor Ronald W. Shephard, was my first thesis advisor. I will judge myself truly successful if I can match the inspiration that he gave to his students, and the respect and admiration he received in return. I also take pleasure in thanking his wife, Mrs. Hilda M. Shephard, for her continuing support and interest in my career.

While he was a fellow student, Professor King-tim Mak, of the University of Illinois at Chicago, first introduced me to the field of Production Theory. My many conversations with him have always been stimulating and rewarding. In addition, his editorial suggestions have helped to greatly improve this thesis from earlier drafts.

Further thanks are due to Professor Ernest Koenigsberg for serving on my committee.

It is impossible to acknowledge all of my friends and colleagues who have helped me to reach this goal. I trust they will accept my humble thanks for I have learned much from each and every one of them. In particular, special thanks go to the following people. Betsy Greenberg provided emotional support and editorial suggestions. Dr. Joerg Boysen listened intently and showed genuine interest. Hyo-sook Lee, Gail Vawter, and Noreen Revelli have advised and supported me throughout. Mariko Kubik and Dr. Ramesh Guttalu did outstanding jobs assisting in the preparation of this manuscript.

Lastly, this acknowledgement would be incomplete if I did not mention Professor Louis J. Billera, of Cornell University, whose unflagging confidence in me allowed me to fulfill my hope of pursuing a graduate career.

Abstract

A theoretical model of a general production process is constructed. A production process is regarded as a network of jointly operating, interrelated activities which use system exogenous inputs of goods and services to produce outputs. The production model displays explicitly the intermediate product transfers between activities and incorporates the time-varying aspects of production directly. The primitive elements which are taken to be common to all production processes are the activity production functions and the flows of products, goods and services. To enhance clarity and rigor, the model is developed axiomatically, i.e., properties on the primitive elements which are conjectured to be true in order to facilitate the theory are identified.

The general model extends previous axiomatic production models ~~of production~~ used in economic theory. Specifically, laws of production and *Shephard's Duality Theorem* are proved using ~~the~~ axioms of the general model. Moreover, the general model provides guidelines as to what entails a satisfactory model of production so that it may be used ~~utilized~~ to study models and solutions of specific production planning problems. To illustrate, the general model is used to systematically analyze a heuristic solution proposed by Leachman and Boysen [1983] for the problem of *multi-project resource-use planning* and to show how their approach can be extended and improved.

TABLE OF CONTENTS

1. INTRODUCTION

In production planning, a decision-maker identifies objectives, available choices, and the result of each choice. Once the particular problem has been described, a model is then formulated as a means of accomplishing the objectives. This includes modeling the result of each choice available to the decision-maker. In production planning, a model of this choice-result relationship is, either indirectly or directly, a model of the actual production process. Thus, modeling the actual production process is a necessary and essential aspect of production planning.

Because the production planner has to develop a model which is tractable for analysis, he often makes certain assumptions which are not specifically identified or adequately justified. Hence, the development of a model for a particular production process is primarily an *art*. To elevate the model-building to a more scientific level would require the modeler to identify *all* of the assumptions taken, either implicitly or explicitly, and justify each one. Therein lies a fundamental problem confronting research in the area of production planning: how does one know what were the implicit and explicit assumptions taken? There are no general guidelines to help the production planner identify the assumptions, and worse, there is no mechanism to systematically analyze the limitations such assumptions necessarily impose.

Another fundamental problem confronting research in the area of production planning is the analysis, or lack thereof, of heuristic solutions. A rigorous analysis of a heuristic solution would be one which provides a rational basis or logical foundation for the proposed methodology. This is impossible unless all of the assumptions of the original model are identified.

As a first step towards elevating the art of production modeling and the analysis of heuristic solutions to a science, we construct a theoretical model of a general production process (Chapter 2). A production process is regarded as a network of jointly operating, interrelated production activities which use system exogenous inputs of goods and services in production to produce final outputs. The production model displays *explicitly* the intermediate product transfers between activities. We identify the primitive elements which we take to be common to

all production processes. They are the activity production functions and the flows of goods and services through time. The model incorporates the time-varying aspects of a production process directly. An *explicit* description of how input is transformed into output is presented. Extensions and limitations of the present model are also discussed.

To enhance clarity and rigor, we develop the model *axiomatically*. That is, properties on the primitive elements which we take to be true in order to facilitate the theory are identified. The chief reasons for axiomatizing the theoretical model are (1) it discloses many of the hidden assumptions, (2) it displays the structure of the theory, (3) the key concepts and hypotheses are identified, (4) the consequences of changes in the foundations are better realized, and (5) the shortcomings of the theory can be spotted and corrected.[1]

Since a specific model of a production process necessarily imposes assumptions on the primitive elements, the general model facilitates the identification of the assumptions. To illustrate, we describe in our framework the production models implicit in the production planning techniques of Material Requirements Planning (MRP) and the ordinary Critical Path Method (CPM). These descriptions will clearly reveal the implicit assumptions about the production processes made by production planners who use such techniques.

Once the implicit assumptions on the primitive elements have been identified, it is possible to analyze any proposed heuristic solution offered to solve a particular problem. To illustrate, we use the general model as a tool to provide a systematic analysis of a heuristic solution proposed by Leachman and Boysen [1983] for the problem of *multi-project resource-use planning* for a multi-project production system (Chapter 4). The problem is to determine explicit resource allocations through time to projects to insure that schedules are met. Our systematic analysis not only provides a logical foundation for their approach but more importantly shows how their approach can be extended and improved. The analyses carried out in this chapter illustrate the value of using a general conceptual framework of a production system to evaluate proposed heuristic solutions to production planning problems.

[1] Adapted from Bunge [1973].

Theoretical models of "general" production processes have been developed before. Recognizing the benefits of axiomatization, Shephard [1970a] developed an axiomatic description of a steady-state production technology.[2] The primitive element in his model was the correspondence which modeled the input-to-output relationship. The axioms were therefore imposed on the correspondences. Realizing that a steady-state framework did not incorporate the dynamic aspects of production directly, Shephard and Fare [1980] extended the framework to model dynamic production systems. Again, the primitive element was the correspondence which modeled the input-to- output relationship. Functions of time were taken to model the flows of goods and services. The axiomatic description for this model was virtually the same as the steady-state model except that certain mathematical axioms were employed to facilitate the theory. Shephard et. al. [1977] developed the first network model of production and later presented an axiomatic description of this model in 1981.[3]

The axiomatic models of production constructed in the past are not useful for the development or evaluation of dynamic production planning models. However, these models have proved to be useful for the development of steady-state cost and production functions.[4] More importantly, perhaps, the previous axiomatic models enable one to prove laws of production. By proving that laws of production hold from the axiomatic description, the question of the validity of the laws is reduced to the question of whether the axioms in the axiomatic framework are appropriate. We continue this worthwhile task by proving, in our general setting, two variants of the *Law of Diminishing Returns* as formulated by Shephard.[5] In addition, we discuss *technical efficiency* and provide two different proofs of *Shephard's Duality Theorem* which better explains this famous theorem (Chapter 3).

Since our model explicitly defines the correspondence which models the input-to-output relationship, our model is more descriptive of the production process than the past axiomatic

[2] See also Shephard [1953], [1970b].
[3] See also Hackman and Shephard [1983].
[4] See, for example, Hanoch and Rothschild [1972].
[5] See Shephard and Fare [1980].

frameworks mentioned above. In addition, since the choice of primitive elements dictates the nature of the axiomatic description, our axiomatic description is completely different from past frameworks. Some of the axioms which were taken in the earlier frameworks were completely *analytical,* i.e, it would not be possible to verify their validity by experimentation. We believe our axiomatic description can be verified through experimentation and is easier to justify.

In summary, the general framework of production introduced here provides guidelines as to what entails a satisfactory model of production. The framework is "general" enough to extend previous axiomatic models useful for developing steady-state cost and production functions and for understanding laws of production in economic theory. Moreover, it is "general" enough to study models and solutions of specific production planning problems.

2. THE GENERAL MODEL

In this chapter, we develop a general model of a production process. Section 2.1 presents the framework of the model. Section 2.2 develops the model axiomatically. Section 2.3 illustrates the generality of the model. Section 2.4 discusses the limitations of the present model and describes how the model could be adapted to fit more specific cases.

2.1. The Framework of the Model

In Section 2.1.1, we present a conceptual framework common to all network models of production.[1] In Section 2.1.2, we describe how we choose to model the flows of goods and services. Sections 2.1.3 and 2.1.4 develop the model of the transformation of input to output at the activity and network levels.

2.1.1. Production Networks: A Conceptual Framework

A production system is modeled as a *directed network*, the nodes of which represent *primitive production activities*. (An example of a Production Network is shown in Figure (2-1).) Primitive production activities are those within which the intermediate product transfers need not be considered for the purposes at hand. The nodes are connected by directed arcs to indicate possible transfers of intermediate and final products. (Cycles are permitted.) System exogenous inputs such as labor services, machine and facility services, energy and fuels, etc., are treated as transfers from an initial node A_0. For a system with N producing activities, final outputs are taken as delivered to node A_{N+1}. Thus, a production system is regarded as a jointly operating, finite number of interrelated primitive production activities $A_1, A_2, \ldots, A_N$ which use system exogenous inputs of goods and services in production to produce final outputs.

Note that the production model displays *explicitly* the intermediate product transfers. This display is essential for dynamic models of production since final output evolves as the evolutionary flow of intermediate products to final products.

[1] See, for example, Shephard et. al. [1977], Shephard [1981], and Hackman and Shephard [1983].

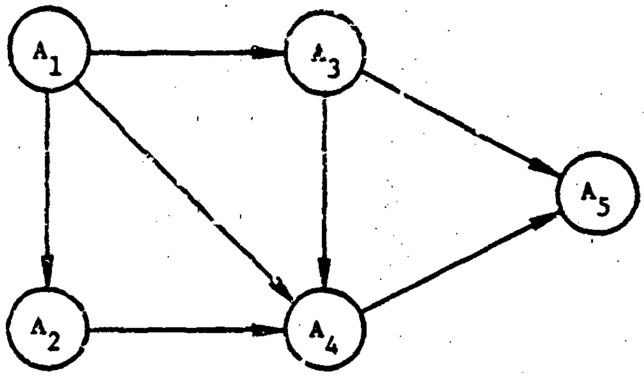


FIGURE (2-1)

EXAMPLE OF A PRODUCTION NETWORK

2.1.2. Modeling the Flows of Goods and Services

The flows of goods and services (input and output) should have a truly dynamic character. A flow, therefore, will be taken to be an element of an appropriate subset of the set of nonnegative functions defined on the nonnegative part of the real line. Each flow will be referred to as a *time-rate history*.[2]

There are two fundamental types of flows. The first and more common type, called a *continuous flow*, is one for which $x(t)$ represents the rate--quantity per unit time--at time t. The second type, called an *event-based flow*, is one for which $x(t)$ is a numerical representation of an event at time t. One example of an event-based flow, suitable for project-oriented production systems (see Section 2.1) is when

$$x(\tau) = \begin{cases} 1 & \text{if } \tau \text{ is the project completion time} \\ 0 & \text{otherwise.} \end{cases}$$

Another example, suitable for batch transfers of intermediate products, is when $x(\tau)$ indicates the quantity transferred at time τ.

2.1.3. The Activity's Dynamic Production Correspondence

How we choose to model the relationship of input into output at the activity level is described in this section. The description of the input-output relationship at the network level is presented in the next section.

To produce output, each activity utilizes system exogenous inputs and intermediate products. The outputs may be intermediate products used as inputs by other activities, or final products, or mixtures of both as in the case of spare parts.

As notation, let

[2] Axioms for the flows of goods and services are presented in Section (2.2.1).

$x_{0i} = (x_{0i}^1, \ldots, x_{0i}^n)$ denote a vector of n time-rate histories of system exogenous inputs allocated to the i^{th} activity, $i=1,2,\ldots,N$,

$V_{ij} = (V_{ij}^1, \ldots, V_{ij}^m)$ denote a vector of m time-rate histories of transfers of outputs from the i^{th} activity to the j^{th} activity, $i=1,2,\ldots,N$, $j=1,2,\ldots,N+1$,

$V_i = (V_i^1, \ldots, V_i^m)$ denote a vector of m time-rate histories of outputs of the i^{th} activity, $i=1,2,\ldots,N$.

Note that in the foregoing representation constants n and m were taken for convenience. That is, a component of x_{0i}, V_{ij} or V_i may be zero.[3]

Abstractly, the *activity's dynamic production correspondence*, denoted by L_i, is taken as a correspondence (set-valued mapping) $V_i \to L_i(V_i)$ where informally one interprets the statement that

$$\left(x_{0i}, \sum_{j=1}^{N} V_{ji}\right) \in L_i(V_i)$$

to mean that the i^{th} activity A_i may produce V_i if allocated x_{0i} as system exogenous input (over time) and $\sum_{j=1}^{N} V_{ji}$ as intermediate product transfers from other activities (over time).

To define the correspondences L_i more formally, one needs to differentiate between the concepts of *allocation* and *application*. Since an activity may be allowed to dispose or store its inputs, what is allocated to the activity as input $\left(x_{0i}, \sum_{j=1}^{N} V_{ji}\right)$ may not be what is actually applied, or consumed, as input into the production process. As notation, let

$y_i = (y_i^1, \ldots, y_i^n)$ denote a vector of n time-rate histories of system exogenous inputs applied into the production process of A_i, $i=1,2,\ldots,N$,

$W_i = (W_i^1, \ldots, W_i^m)$ denote a vector of m time-rate histories of intermediate product transfer inputs applied into the production process of A_i, $i=1,2,\ldots,N$,

[3] The term "zero" here refers to the function $x: R_+ \to R_+$ such that $x(t)=0$, $\forall t \in R_+$. $R_+ = \{t \mid t \geq 0\}$.

$T_i = (T_i^1, \ldots, T_i^n, T_i^{n+1}, \ldots, T_i^{n+m})$ denote a vector of $n+m$ time-rate histories of disposal for both input and output for A_i, $i=0,1,2,\ldots,N,N+1$.

Given a particular allocation of input $\left(x_{0i}, \sum_{j=1}^{N} V_{ji}\right)$ those choices for applications of input y_i, W_i and disposal T_i which are feasible are those for which the following inventory balance constraints are satisfied:[4]

if the j^{th} system exogenous input is activity storable,

$$0 \leqslant c_i^j + \int_0^t \{x_{0i}^j - T_i^j - y_i^j\} d\mu \leqslant C_i^j, \quad \forall t \in R_+ \tag{2.1}$$

if the k^{th} product is activity storable,

$$0 \leqslant b_i^k + \int_0^t \{\sum_j V_{ji}^k - T_i^{n+k} - W_i^k\} d\mu \leqslant B_i^k, \quad \forall t \in R_+ \tag{2.2}$$

if the j^{th} system exogenous input is not activity storable,

$$0 = x_{0i}^j - T_i^j - y_i^j \tag{2.3}$$

if the k^{th} product is not activity storable,

$$0 = \sum_j V_{ji}^k - T_i^{n+k} - W_i^k \tag{2.4}$$

where $c_i^j, b_i^k \in R_+$ represent initial stocks, if any, and $C_i^j, B_i^k \in R_+ \cup \{\infty\}$ represent the constant capacity levels. It is understood that T_i^j (T_i^k) is zero if the j^{th} system exogenous input (k^{th} product) is not disposable.

The model at the activity level assumes the existence of a *production function*, denoted by F_i, which takes a vector of inputs applied to the production process (y_i, W_i) into *realized* vector

[4] The measure μ will be defined precisely in Axiom 1 next section. The set on which we are integrating is $[0,t]$.

(of dimension m) of outputs of the activity obtained through production, $F_i(y_i, W_i)$.[5] Since activities may be allowed to dispose or store products produced $F_i(y_i, W_i)$ need not be equal to V_i. By the expression "$F_i(y_i, W_i)$ is enough to support output level V_i" we mean that the following inventory balance constraints are satisfied:

if the k^{th} product is activity storable,

$$0 \leq c_i^k + \int_0^t \{F_i^k(y_i, W_i) - T_i^{m+k} - V_i^k\} d\mu \leq B_i^k, \quad \forall t \in R_+, \tag{2.5}$$

if the k^{th} product is not activity storable,

$$0 = F_i^k(y_i, W_i) - T_i^{m+k} - V_i^k . \tag{2.6}$$

Thus, to say that $\left[x_{0i}, \sum_{j=1}^{N} V_{ji}\right] \in L_i(V_i)$ we mean that there exist an application vector (y_i, W_i) and disposal vector T_i such that (2.1)-(2.6) are satisfied. We now turn to describing the input-output transformation at the network level.

2.1.4. The Network Dynamic Production Correspondence

Let $u = (u^1, \ldots, u^m)$ denote a vector of *final output* time-rate histories. The network dynamic production correspondence, denoted by LN, is a correspondence $u \rightarrow \mathrm{LN}(u)$ which, loosely described, is the set of all vectors of system exogenous input rate histories x that when appropriately allocated to the activities may produce the vector u of final output rate histories.

To be specific, $x \in \mathrm{LN}(u)$ means that we can find allocations of system exogenous inputs to the activities, x_{0i}'s, and allocations of intermediate products to activities obtained from other activities, the V_{ij}'s, such that the following inventory balance constraints are satisfied:

On the input side,

if the j^{th} system exogenous input is system storable,

[5] Axioms for the Activity Production Functions are presented in Section (2.2.2).

$$0 \leqslant b^j + \int_0^t \{x^j - T_0^j - \sum_i x_{0i}^j\} d\mu \leqslant B^j, \quad \forall t \in R_+ \tag{2.7}$$

if the j^{th} system exogenous input is not system storable,

$$0 = x^j - T_0^j - \sum_i x_{0i}^j \tag{2.8}$$

where $b^j \in R_+$ is the initial stock, if any, and $B^j \in R_+ \cup \{\infty\}$ represents the constant capacity. It is understood that $T_0^j = 0$ if the j^{th} system exogenous input is not system disposable.

On the output side,

if the k^{th} product is system storable,

$$0 \leqslant b_{N+1}^k + \int_0^t \{\sum_i V_{i,N+1}^k - T_{N+1}^{n+k} - u^k\} d\mu \leqslant B_{N+1}^k, \quad \forall t \in R_+, \tag{2.9}$$

if the k^{th} product is not system disposable,

$$0 = \sum_i V_{i,N+1}^k - T_{N+1}^{n+k} - u^k \tag{2.10}$$

where $b_{N+1}^k \in R_+$ is the initial stock, if any, and $B_{N+1}^k \in R_+ \cup \{\infty\}$ represents the constant capacity. It is understood that $T_{N+1}^{n+k} = 0$ if the k^{th} product is not system disposable.

Finally, one needs to insure that the individual activities can produce what is required of them, so one adds

$$\left(x_{0i}, \sum_j V_{ji}\right) \in L_i(V_i), \quad i = 1, 2, \ldots, N. \tag{2.11}$$

Two important comments are in order with respect to the definition of the correspondence LN. First, we have tacitly assumed, for each i,

$$V_i = \sum_{j=1}^{N+1} V_{ij} \ . \tag{2.12}$$

This convenient assumption is not restrictive because:

(1) Storage and disposal of intermediate products are already allowed at the activity level,

(2) If the production system is such that it is more appropriate to dispose or store intermediate products at the system level--for example, products are warehoused--then the acts of disposal and storage may be simply modeled as a separate activity. If the acts of storing and disposing were costly, then modeling such acts as activities would be appropriate.

The second comment concerns the expression $\sum_{j=1}^{N} V_{ji}$ (the vector sum of intermediate product transfer inputs from all activities into activity i). It is clear from this expression that our model does not incorporate transfer or shipment lags. To incorporate this time lag, one could define $V_{ji}(t)$ as what A_j *sends* A_i at time t and introduce $V_{ji}^*(t)$ as what A_i *receives* from A_j at time t. For example, if there were a constant time lag l_{ji} for shipment then

$$V_{ji}^*(t) = \begin{cases} V_{ji}(t - l_{ji}) & \text{if } t \geq l_{ji} \\ 0 & \text{if } 0 \leq t \leq l_{ji} . \end{cases}$$

Finally, we make a useful definition. If $x \in \mathrm{LN}(u)$, there are many possible collections of flows x_{0i}'s, V_{ij}'s, y_i's W_i's, and T_i's of goods and services which satisfy (2.1)-(2.11). By the expression "a feasible flow for input x to support output level u", we mean one such collection of flows. On occasion we will simply say "a feasible flow to support output level u" if reference to a specific x is not required and a "feasible flow" if reference to a specific u is not required.

2.2. An Axiomatic Presentation of the General Model

In Section 2.1, we provided the framework of the general model. We first introduced production networks as a conceptual model. Then we modeled the flows of goods and services so that they would be truly dynamic. Finally, we described how input is transformed into output.

From the discussion in Section 2.1, it is clear that two types of *primitive elements* comprise the model:

(1) the flows of goods and services, and

(2) the activity production functions.

The purpose of this section is to define precisely these elements. As mentioned in the introduction, the best way of being precise is through the axiomatic method. In Section 2.2.1, we develop and justify the axioms taken for the flows of goods and services. In Section 2.2.2, we develop and justify the axioms taken for the activity production functions.

We remind the reader that it is our attitude that an axiom is not an a priori truth but rather a scientific hypothesis conjectured in order to facilitate a theory. The justification of each axiom rests mainly on the assertion that the property imposed is a property one would expect to observe in the future or one has observed in the past.

2.2.1. Axioms for the Flows of Goods and Services

2.2.1.1. Axiom 1: The Underlying Space of Flows of Goods and Services

It was argued in Section 2.1 that to represent the truly dynamic character of production, a flow of a good or service should be modeled as a nonnegative function of time. That is, if z represents a flow then

$$z: R_+ \rightarrow R_+$$

where $R_+ = \{t : t \geq 0\}$ models the time axis.

The mathematical operation of integration was employed to define the inventory balance constraints (2.1)-(2.11) presented in Section 2.1. To formally integrate a flow, it requires one to first define a measurable space with respect to which the flow is measurable. If $B(R_+)$ denotes the Lebesgue σ-field of R restricted to R_+, then we take $(R_+, B(R_+))$ to be the measurable space. Second, one must select an appropriate measure μ on $B(R_+)$.

The measure μ must account for *both* continuous and event-based flows. Since a continuous flow measures the quantity per unit time continuously through time, Lebesgue measure, denoted by λ, is a suitable measure which may be used to integrate a continuous flow. However, Lebesgue measure is *not* suitable for event-based flows. Events, as we choose to think of them, do not occur continuously through time. That is, the set on which an event-based flow is positive has Lebesgue measure equal to zero. A suitable measure which may be used to integrate an event-based flow is a *counting measure*, denoted by ν, of the following type:

A countably infinite index set $T = \{t_k\}_1^\infty \subset B(R_+)$ is assumed to exist for which if $B \in B(R_+)$,

$$\nu(B) = \begin{cases} |B \cap T| & \text{if } |B \cap T| < \infty \\ \infty & \text{otherwise} \end{cases}$$

where $|A|$ denotes the cardinality of the set A.

Each point in T indicates a *possible* time of an event. So, to account for both continuous and event-based flows the measure μ is taken to be of the form $\lambda + \nu$. We are now ready to state Axiom 1.

Axiom 1

If z represents a flow of *a* good or service, then $z \in L_+^\infty(R_+, B(R_+), \lambda + \nu)$ where L_+^∞ denotes the nonnegative orthant of L^∞. If T denotes the index set associated with ν, then T is assumed to satisfy the following properties:

(1) $0 = t_0 < t_1 < t_2 < \cdots < t_n < t_{n+1} < \cdots$

(2) $\inf_{k \geq 1} \{t_k - t_{k-1}\} > 0$

(3) $\sup_{k \geq 1} \{t_k - t_{k-l}\} < \infty$.

We make 3 comments about Axiom 1:

(A) The restrictions on T were imposed to allow us to view $I_k = (t_{k-1}, t_k)$ as a *period* in production planning, the points in T as *time-grid points*, and T as the *time grid*. The assumption of a time grid T satisfying properties (1)-(3) above is always assumed in discrete-time planning and control.

(B) Since a flow is defined up to sets of measure zero, it is understood that a continuous flow is constrained to be one for which $z(t) = 0$ if $t \in T$ and an event-based flow is one for which $z(t) = 0$ if $t \notin T$.

(C) Axiom 1 states that a flow is initially constrained to be a function which is (i) nonnegative, (ii) measurable, and (iii) essentially bounded. The set of functions satisfying these three properties is large. It includes functions which are not flows of goods and services one expects to occur in production. Therefore, additional constraints in the following axioms need to be imposed to define what constitutes an acceptable flow. The constraints to be imposed will apply to each *flow type*. By the expression "flow type", we refer to a particular class of flows. For example, one of the flow types is the class of flows corresponding to the allocation of the j^{th} input to the i^{th} activity which we have denoted in Section 2.1 by the symbol x_{0i}^j.

2.2.1.2. Axiom 2: Limiting the Shape of the Flows of Goods and Services

Each flow is (essentially) bounded (in norm) over the infinite horizon and hence in each period. We further insist that a flow's bound in a period is bounded by a function of the cumulative amount of the flow in that period.

The thrust of Axiom 2 is to limit flows which are sharply "peaked." That is, Axiom 2 in effect limits the deviation between the bound of a flow in a period, $\|z \cdot 1_{I_k}\|_\infty$, and the mean value of the flow in the period, $\dfrac{\int_{I_k} z d\mu}{\mu(I_k)}$. An example of a flow which is sharply peaked is graphically depicted in Figure (2-2).

The restriction to be imposed in Axiom 2 is made *parametrically*. The parameters are not constants but continuous non-decreasing nonnegative scalar valued functions on R_+.

Axiom 2

For the i^{th} flow type, there exists families of parameters $\{g_k^i\}_{k=1}^\infty$, $\{h_k^i\}_{k=1}^\infty$, so that if z is a flow of this type then for all k

(1) if $\int_{I_k} z d\mu \leqslant A$ then $\|z \cdot 1_{I_k}\|_\infty \leqslant g_k^i(A)$

(2) if $\int_{\{t_k\}} z d\mu \leqslant A$ then $\|z \cdot 1_{\{t_k\}}\|_\infty \leqslant h_k^i(A)$[1]

In discrete-time planning and control, the assumption almost always taken is that each flow is a continuous flow which is a *step-function associated with the time grid*. That is, a flow is assumed to be constant on each period I_k. Since our time grid T does not change, we will refer to these functions as step-functions. The justification for the structure of the property imposed in Axiom 2 is that if we properly select the parameters $\{g_k^i\}_1^\infty$, $\{h_k^i\}_1^\infty$ then Axiom 2 may be used to limit a flow to the set of step-functions. Moreover, we may select parameters in such a way so as to restrict the range of a flow.

Proposition (2.1)

[1] Of course, this reduces to the statement that $x(t_k) \leqslant A \rightarrow x(t_k) \leqslant h_k(A)$.

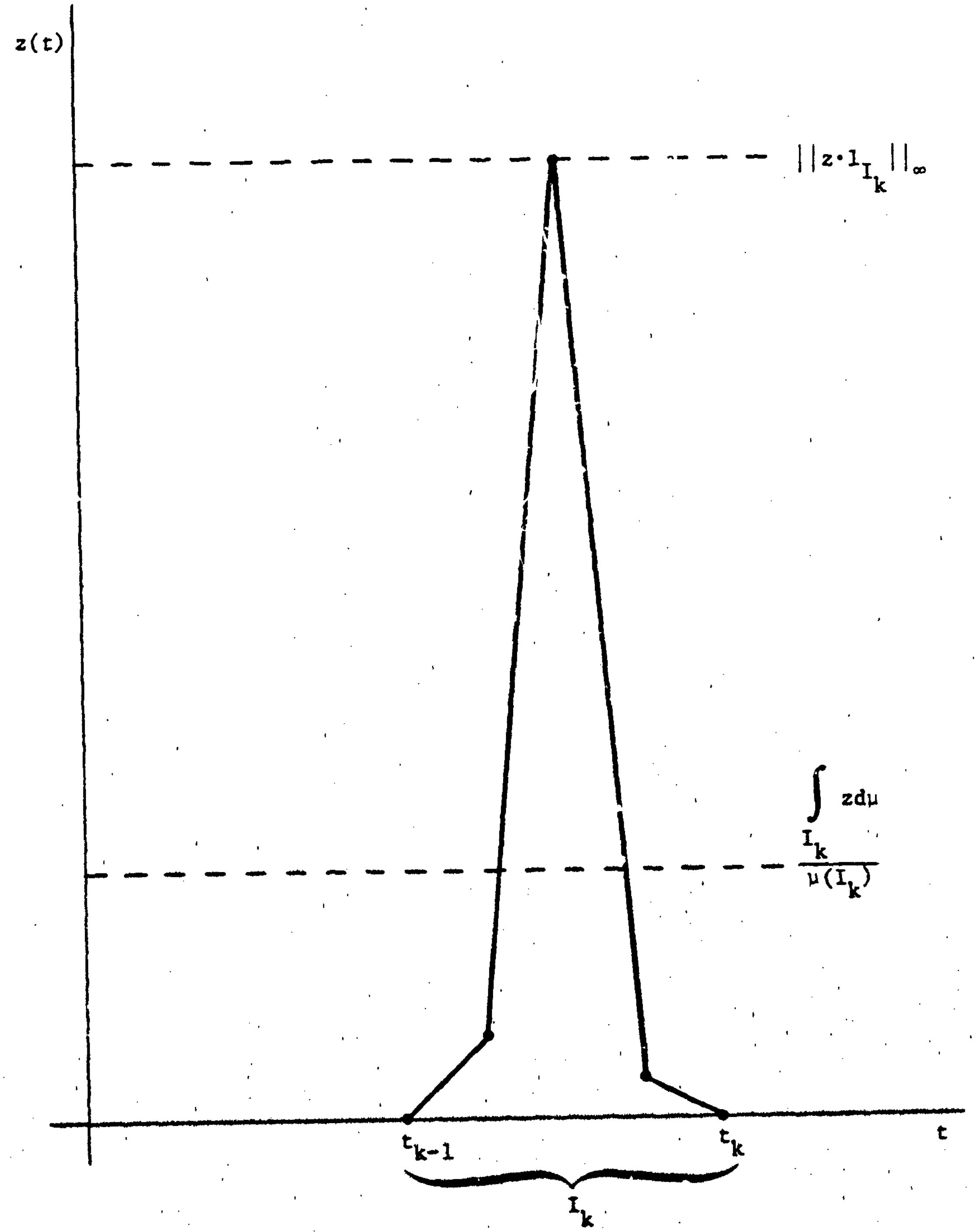


FIGURE (2-2)

EXAMPLE OF A SHARPLY PEAKED FLOW

Let S be a non-empty closed subset of R_+. Then there exists parameters $\{g_k\}_1^\infty$, $\{h_k\}_1^\infty$ such that if z is a flow satisfying Axiom 2 (with these parameters) then z must be a step function whose range is contained in S.

Proof of Proposition (2.1)

Since S^C is open in R, write it as a countable, disjoint union of intervals $I_n = (a_n, b_n)$, i.e., $S^C = \bigcup_{n=1}^{\infty} I_n$.[2]

For each k, define g_k and h_k as follows:

$$g_k(A) = \begin{cases} \dfrac{A}{\lambda(I_k)} & \text{if } \dfrac{A}{\lambda(I_k)} \in S \\[2ex] a_n + \dfrac{\left(\dfrac{A}{\lambda(I_k)} - a_n\right)^2}{b_n - a_n} & \text{if } \lambda(I_k)a_n < A < \lambda(I_k)b_n,\ a_n \geqslant 0,\ b_n < \infty \\[2ex] \dfrac{\left(\dfrac{A}{\lambda(I_k)}\right)^2}{b_n} & \text{if } \lambda(I_k)a_n < A < \lambda(I_k)b_n,\ a_n < 0,\ b_n < \infty \\[2ex] \tfrac{1}{2}\left(\dfrac{A}{\lambda(I_k)} - a_n\right) & \text{if } \lambda(I_k)a_n < A < \lambda(I_k)b_n,\ b_n = \infty . \end{cases} \tag{2.13}$$

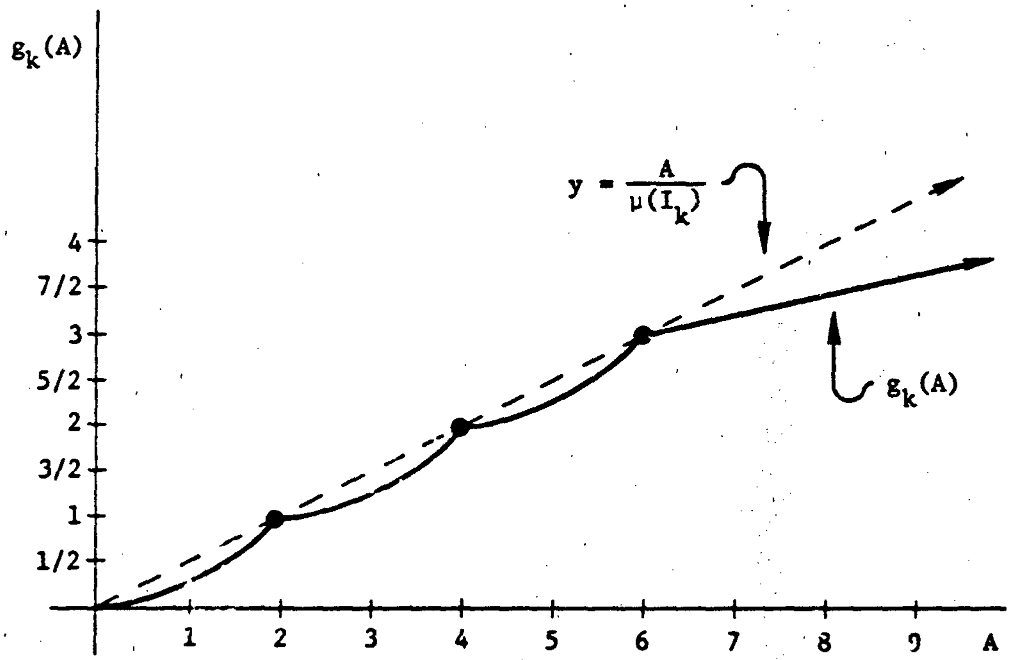
$$h_k(A) = \tfrac{1}{2}A . \tag{2.14}$$

Figure (2-3) exhibits a pictorial representation of the graph of a typical g_k. Here, the set $S = \{2,4,6\}$. Inspection of (2.13) and (2.14) shows that each g_k and h_k are acceptable parameters.

Since the nonnegative a_n's and b_n's necessarily lie in S, it is a simple matter to check that

$$\left\{A \in R_+ \mid g_k(\lambda(I_k)A) < A\right\} = R_+ \tag{2.15}$$

[2] Royden [1968], p. 39.



$\mu(I_k) = 2$

$S = \{2,4,6\}$

FIGURE (2-3)

DEMONSTRATION OF PARAMETERIZATION FOR STEP-FUNCTION CASE

$$\left\{A \in R_+ \mid g_k(\lambda(I_k)A) = A\right\} = S \tag{2.16}$$

Let z be a flow satisfying (1) and (2) of Axiom 2. By (2) and our choice of h_k's, we have that, for each k,

$$z(t_k) \leq h_k(z(t_k)) = \tfrac{1}{2}z(t_k)$$

which implies that $z(t_k) = 0$ for each k. Hence, z must be a continuous flow.

By (1) and (2.15), we also have that, for each k, z must satisfy the following inequalities:

$$\begin{aligned}
\int_{I_k} z\,d\mu &\leq \|z \cdot 1_{I_k}\|_\infty \lambda(I_k) \\
&\leq g_k\Big(\int_{I_k} z\,d\mu\Big)\lambda(I_k) \quad \text{by (1)} \\
&\leq \int_{I_k} z\,d\mu \qquad \text{by (2.15)}.
\end{aligned}$$

Hence, the above inequalities must be equalities which implies that

$$g_k\Big(\int_{I_k} z\,d\mu\Big) = \frac{\int_{I_k} z\,d\mu}{\lambda(I_k)} \quad \text{for each } k \text{, and} \tag{2.17}$$

$$z(t) = \frac{\int_{I_k} z\,d\mu}{\lambda(I_k)} \quad \text{if } t \in I_k. \tag{2.18}$$

In view of (2.16) and (2.17), it is immediate that the constant

$$\frac{\int_{I_k} z\,d\mu}{\lambda(I_k)} \in S.$$

By (2.18), the result follows.■

2.2.1.3. Axiom 3: Existence of a Lower Bound

Axiom 3 essentially states that if a flow is positive, then it must be some minimal value.

Axiom 3

For the i^{th} flow type, there exists an $\epsilon_i > 0$ such that if z is a flow of this type then

$$\mu\{0 < z < \epsilon_i\} = 0 .$$

In production planning problems, usually only the assumption of nonnegativity on the flows is imposed. This assumption is made in order to facilitate the implementation of algorithms used to solve the formulation of the problem. We believe, however, that flows one would observe or have observed in production satisfy Axiom 3.

2.2.1.4. Axiom 4: Limitation on Set-up times

Given a flow z, a *set-up* time for z, loosely worded, is a time τ at which z is either "starting up again" or "stopping." If $z \in L_+^\infty(R_+, B(R_+), \mu)$, then we define the set of set-up times for z to be

$$S(z) = \Big\{\tau \in R_+ \mid \text{if I is an open interval, } \tau \in I \subset R_+, \text{then}$$

$$\mu\{I \cap (z > 0)\} > 0 \text{ , and } \mu\{I \cap (z = 0)\} > 0\Big\} . \tag{2.19}$$

Axiom 4 insists that the time between any two set-up times for a flow is at least some minimal value.

Axiom 4

For the i^{th} flow type, there exists a $\delta_i > 0$ so that if z is a flow of this type then

$$\inf_{\tau,\tau' \in S(z)} |\tau' - \tau| \geq \delta_i .$$

Before we proceed to present the axioms taken for the activity production functions, it will be useful to introduce some notation. We let $L_+^\infty(\mu)$ denote $L_+^\infty(R_+, B(R_+), \mu)$. We let the generic symbol L_i denote that subset of $L_+^\infty(\mu)$ which satisfies Axioms 2-4 for flow type i. When referring to a particular flow type, we will use the symbol which refers to flows of that type (for example, $L_{x_{i_k}}$).

2.2.2. Axioms for the Activity Production Functions

2.2.2.1. Axiom 5: Closure of the Domain

Let $L_{y_i} = L_{y_i 1} \times \cdots \times L_{y_i a}$, $L_{W_i} = L_{W_i 1} \times \cdots \times L_{W_i m}$. If D_i denotes the domain of the function F_i, then $D_i \subset L_{y_i} \times L_{W_i}$. If there are additional constraints linking the domains of the inputs applied in production then D_i may be a proper subset of $L_{y_i} \times L_{W_i}$.[3] For example, in the Leontief-type input-output models of production the inputs applied into production, if non zero, are assumed to be proportional. Another example is when an activity utilizes one machine to produce several similar types of products. In this example, the exogenous inputs applied into production include the rate of machine hours applied to each type of product. The functions which define such rates are linked in that no two of them are positive at the same time.

Axiom 5

[3] Production systems exist where constraints are imposed on the applications of inputs belonging to *different* activities. We choose not to incorporate this feature in our general model. Such cases will have to be examined on an individual basis.

Endow L_k with the relative weak-star topology for each k.[4] It is assumed that D_i is closed in the product topology on $L_{y_i} \times L_{w_i}$.

It is necessary for mathematical reasons that the domain for each production function is closed. It is not possible to examine each instance of additional constraints linking the domains of the inputs applied in production and show closure in each case. Hence, Axiom 5 is imposed. However, it will be shown that if $D_i = L_{y_i} \times L_{w_i}$ then D_i is closed in the weak-star product topology on $(L^\infty)^{n+m}$. (Proposition (3.4.2), Appendix, Ch. 3). Furthermore, the domains associated with the two examples given above are also shown to satisfy Axiom 5 (Proposition (3.4.6), Appendix, Ch. 3). The purpose of examining these examples and showing that Axiom 5 holds in each case is to give a plausible basis for accepting Axiom 5.

2.2.2.2. Axiom 6: Null Activities are Excluded

An obvious property to impose on each production function is to insure that each activity is capable of producing at least one product.

Axiom 6

For all A_i, F_i on D_i is non-trivial. That is, $\exists k$ and a $(y_i, W_i) \in D_i$ s.t. $\|F_i^k(y_i, W_i)\|_\infty > 0$.

2.2.2.3. Axiom 7: Existence of System Essential Inputs

A subset of system exogenous inputs is called *essential* if when none of these inputs are applied in production then no output of any kind is possible. Typically, in most production systems labor services is an example of a set of essential inputs.

[4] A basic open neighborhood N of zero in this topology has the following form: $\exists p^1, \ldots, p^n \in L^1$ and an $\epsilon > 0$ such that $N = \{z \in L^\infty \mid |\int (p^i \cdot z) d\mu| < \epsilon,\ \forall i\}$. N is usually denoted by $N(p^1, \ldots, p^n; \epsilon)$.

Axiom 7

There exists a set E, $\varnothing \neq E \subset \{1,2,\ldots,n\}$, $|E| < n$, such that for each A_i if $(y_i, W_i) \in D_i$ with $y_i^j = 0$ for $j \in E$ then $F_i(y_i, W_i) = 0$.

Any set satisfying the above property will be called an *essential set of system exogenous inputs.*

2.2.2.4. Axiom 8: Boundedness of the Production Function

If the rates of application associated with an essential set of system exogenous inputs applied into an activity's production process were bounded, then we maintain that the output rates realized through production would be bounded *regardless* of the magnitude of the rates of application of the intermediate product input. This axiom suggests itself from experience.

Axiom 8

Let E be an essential set of system exogenous inputs. For $C \in R_+$, let

$$Y_C = \{y \in L_{y_i} \mid \max_{j \in E} \|y^j\|_\infty \leq C\}.$$

Then for each A_i

$$\sup_{(y,W) \in D_i, y \in Y_C} \left\{ \max_{1 \leq k \leq m} \|F_i^k(y,W)\|_\infty \right\} < \infty, \quad \forall C \in R_+$$

2.2.2.5. Axiom 9: Input-Output Continuity

Let (y_1, W_1), (y_2, W_2) be two input vectors belonging to the domain D_i of F_i. Let $h > 0$. Fix the horizon to the finite interval $[0,h]$. If the difference in the cumulative amount of each input in each period between the two input vectors is sufficiently small, then we maintain that the difference in the cumulative amount of each output over the horizon between the realized output vectors of production must necessarily be small.

Axiom 9

For each A_i and $\forall h \in R_+$, if $(y_1, W_1) \in D_i$, then $\forall \epsilon > 0$, $\exists \delta > 0$ so that if $(y_2, W_2) \in D_i$ satisfies

$$\text{(i)} \quad \max_l \left| \int_{\bar{I}_l \cap [0,h]} (y_1^j - y_2^j) d\mu \right| < \delta, \quad 1 \leqslant j \leqslant n,$$

$$\text{(ii)} \quad \max_l \left| \int_{\bar{I}_l \cap [0,h]} (W_1^k - W_2^k) d\mu \right| < \delta, \quad 1 \leqslant k \leqslant m,$$

then

$$\left| \int_{[0,h]} (F_i^k(y_1, W_1) - F_i^k(y_2, W_2)) d\mu \right| < \epsilon, \quad 1 \leqslant k \leqslant m.$$

There are examples of production functions which do not satisfy Axiom 9. One example is the production function associated with a chemical process which produces a successful reaction only when a certain level or *critical threshold* of input is reached. To incorporate these activity production functions into an framework one may either choose to modify the domain D_i so that levels below the critical threshold no longer belong to the domain or modify the production function itself. We believe that either modification can be made without seriously affecting the model of the activity's production process.

2.2.2.6. Axiom 10: Efficiency of the Production Function

If the vector of intermediate product inputs applied into an activity's production process is fixed, then we maintain that there exists a bound on the cumulative amount of exogenous inputs applied in production before the *activity operates inefficiently*. By the expression "activity operates inefficiently" we mean that an activity could produce at least as much cumulative output while using "less" exogenous input.

The bound on the cumulative amount of exogenous inputs applied into production is assumed to be a function of the vector of intermediate products applied into production. It is also assumed that this function itself is bounded in a manner similar to Axiom 8.

Axiom 10

For each A_i there exists a $g_i : L_{W_i} \times R_+ \rightarrow R_+$ such that $\forall h \in R_+$ if $(y_1, W) \in D_i$ with

$$\max_{1 \leqslant j \leqslant n} \int_{[0,h]} y_1^j d\mu \geqslant g_i(W,h)$$

then there exists a $y_2 \lneq y_1$ such that $(y_2, W) \in D_i$ and such that

$$\int_{[0,h]} F_i^k(y_2, W) d\mu \geqslant \int_{[0,h]} F_i^k(y_1, W) d\mu , \quad 1 \leqslant k \leqslant m .$$

Let $W_C = \{ W \in L_{W_i} \mid \max_{1 \leqslant k \leqslant m} \| W^k \|_\infty \leqslant C \}$. Then each g_i is assumed to satisfy the following property: $\forall h \in R_+$,

$$\sup_{W \in W_C} g_i(W,h) < \infty$$

2.2.2.7. Axiom 11: Past Production Not Affected by Future Inputs

Let (y_1, W), (y_2, W) be two input vectors in the domain D_i such that for some $h > 0$.

$$y_2^j = y_1^j \cdot 1_{[0,h]}, \quad 1 \leqslant j \leqslant n .$$

That is, y_2 is the restriction of y_1 to the horizon $[0,h]$. Since no future input can affect the output already generated up to time h, then we maintain that

$$F_i^k(y_2, W) \cdot 1_{[0,h]} = F_i^k(y_1, W) \cdot 1_{[0,h]}, \quad 1 \leqslant k \leqslant m$$

even though mathematically $(y_2, W) \neq (y_1, W)$.

Axiom 11

For each A_i and $\forall h \in R_+$ if $(y, W) \in D_i$ then

$$F_i^k((y^1 \cdot 1_{[0,h]}, \ldots, y^n \cdot 1_{[0,h]}), W) = F_i^k(y, W) \cdot 1_{[0,h]}, \quad 1 \leqslant k \leqslant m .$$

2.2.2.8. Axiom 12: Cumulative Production Limited in Finite Horizon

We maintain that in a finite amount of time, an activity cannot produce an infinite amount of output *regardless* of the inputs applied.

Axiom 12

For each A_i and $\forall h \in R_+$,

$$\sup_{(y,W)\in D_i} \int_{[0,h]} F_i^k(y,W)\,d\mu < \infty, \quad 1 \leqslant k \leqslant m .$$

We make an important observation about the implication of Axiom 12. If in a finite amount of time only a finite amount of cumulative output may be realized through production, then the cumulative amounts of the allocations and applications of intermediate product input associated with any feasible flow must be finite in a finite horizon. By Axiom 2, this implies that in a finite horizon the flows of the allocations and applications of intermediate product input associated with any feasible flow are uniformly bounded in norm. Thus, a function $B: R_+ \to R_+$ exists such that $B(h)$ is such a bound. The function $B(h)$ is a function of the horizon $[0,h]$ and the parameters used to define Axioms 2 and 12.

2.3. Special Cases of the General Model

In this section, we illustrate the model's generality. In Section 2.3.1, we use the description of the general model to define precisely the Dynamic Linear Activity Analysis Model of Production. In Section 2.3.2, we show how the Traveling Salesman Problem may be embedded into the framework of the general model. In Section 2.3.3, we discuss Material Requirements Planning. Finally, in Section 2.3.4, we discuss single-project production systems.

2.3.1. The Dynamic Linear Activity Analysis Model (DLAAM)

The *Dynamic Linear Activity Analysis Model of Production (DLAAM)* is an extension of the dynamic Leontief input-output models of production.[1] All of these models assume a particular *form* for the domain of the activity production functions and the production functions themselves.

The Activity Production Functions

Each domain D_i is assumed to have the following form: there exist constants a_{ij}, $j = 1,2,\ldots,n$, $\bar{a}_{ik}$, $k = 1,2,\ldots,m$ such that

$$D_i = \{(y_i, W_i) \mid \exists z_i \text{ such that } y_i^j = a_{ij} z_i \,,\; W_i^k = \bar{a}_{ik} z_i\}.$$

In other words, if the applications of inputs are positive, then they must be *proportional*; thus, they may be indexed in terms of one profile, z_i, called the *intensity curve.* The intensity curve z_i is assumed to be a bounded step-function associated with an equal-length period time grid.

The production function F_i is a function of (y_i, W_i). In view of the form of the domain D_i we may write the production function as a function of the intensity curve denoted by $F_i(z_i)$. Each $F_i(z_i)$ is assumed to have the following form: there exists constants c_{ik}, $k = 1,2,\ldots,m$ such that

[1] First developed in Shephard, Al-Ayat, Leachman [1977] and later adapted by Leachman [1982]. See also Leontief [1951], Koopmans [1951], and Morgenstern [1954].

$$F_i^k(z_i) = c_{ik} z_i, \quad \text{for each } k .$$

In other words, the outputs generated are assumed proportional (if positive) and indexed by the same profile as the one used to index the inputs.

Inventory Calculations

In the general model presented in Chapter 2, the inventory of a good or service was constrained to be nonnegative and less then capacity at *all* points in time. DLAAM relaxes this restriction in that the calculations of inventory only occur at the time grid points. For a finite horizon, this relaxation implies that the level sets $\mathrm{LN}(u)$ may be described in terms of a finite set of linear inequalities. (The cumulative amounts of each flow in each period become the variables.) Thus, formulations of production planning problems which assume DLAAM as a model of production permit the use of linear programming.

The general inventory balance constraints imposed on the intermediate product transfers are:

$$\int_{[t_1,t_r]} \sum_i V_i^k(\tau - t_1)\, d\mu \geqslant \int_{[t_1,t_r]} \sum_i W_i^k\, d\mu , \quad \forall_r \geqslant t_1 ,\ 1 \leqslant k \leqslant m . \tag{2.20}$$

A lag of one period is incorporated in (2.20) to insure feasibility since the calculations of the inventory only occur at the time grid points. Note that DLAAM does not explicitly model each V_{ij} but only the sums $V_i = \sum_i V_{ij}$ and $W_i = \sum_j V_{ji}$. This is because DLAAM does not impose any side constraints on the V_{ij}'s. Without side constraints on the V_{ij}'s, these variables become redundant. In fact, most production models do not explicitly mention the V_{ij}'s precisely because they do not impose side constraints on the V_{ij}'s. For a further discussion of DLAAM and its relationship to Leontief-type input-output models of production see Leachman [1984].

2.3.2. The Traveling Salesman Problem

Let H be a directed network with N nodes, $N \geqslant 3$. The traveling salesman problem is to find a minimum-length Hamiltonian cycle, i.e., a cycle passing through each node exactly once. As notation, let $d_{(i,j)} =$ length of arc (i,j). We assume $d_{(i,j)}$ is a positive integer. We let A denote the set of arcs in H.

In this section, we will construct a production system such that, for an appropriate choice of the final output vector u, each feasible flow for u corresponds to a Hamiltonian cycle and vice-versa. *Moreover, what is being produced by this production system is the length of the Hamiltonian cycle.* The development also illustrates the very general ways in which the primitive elements can be defined.

We first proceed to construct the production network G associated with the production system. The nodes in G correspond to the arcs in H; that is, if arc (i,j) exists in H then node $A_{(i,j)}$ exists in G. We also add a "sink" node $A_{(N+1,N+1)}$. There is an arc from $A_{(i,j)}$ to $A_{(k,l)}$ if $(i,j) \neq (k,l)$ and $j = k$. Also, for each $(i,j) \neq (N+1,N+1)$ an arc from $A_{(i,j)}$ to $A_{(N+1,N+1)}$ is added.

To define the production system associated with G, we need to define the limitations on the flow types. We list these below.

(1) The time grid T is the set of whole numbers $\{0,1,2,\cdots\}$.

(2) All flows are event-based.

(3) The range of each flow belongs to $\{0,1\}$.

(4) Except at the sink node, no inventories of any kind are allowed.

(5) No disposal of any kind is allowed.

(6) There is only one exogenous input and only one product.

(7) Attention is restricted to the finite horizon $[0,h]$ where $h = \sum_{(i,j)\in A} d_{(i,j)}$.

The additional constraints imposed on the flows--other than those constraints already imposed by the general model--are listed below:

$$\int_{[0]} \left\{ \sum_{(k,l)\neq(N+1,N+1)} y_{(k,l)} \right\} d\mu = 1 \tag{2.21}$$

$$\int_{[0,h]} \left\{ \sum_{\{(i,l) \mid (i,j),(j,l) \in A\}} V_{(i,j),(j,l)} \right\} d\mu = 1, \quad j \neq N+1 \tag{2.22}$$

$$y_{(i,j)} = 0 \rightarrow V_{(i,j),(N+1,N+1)} = 0 . \tag{2.23}$$

The production functions are defined as follows:

$$F_{(k,l)}(y_{(k,l)}, W_{(k,l)}) = \begin{cases} 0 & \text{if } y_{(k,l)} = 0,\ W_{(k,l)} = 0 \\ 1_{\{d_{(k,l)}\}} & \text{if } y_{(k,l)} \neq 0,\ W_{(k,l)} = 0 \\ 1_{\{d(k,l)\}} + 1_{\{\tau(W_{(k,l)})\}} & \text{if } y_{(k,l)} \neq 0,\ W_{(k,l)} \neq 0 \\ 1_{\{\tau(W_{(k,l)}) + d_{(k,l)}\}} & \text{if } y_{(k,l)} = 0,\ W_{(k,l)} \neq 0 \end{cases} \tag{2.24}$$

where $\tau(W_{(k,l)}) = \min\{n \mid W_{(k,l)}(n) = 1\}$. A routine check shows that all flow types satisfy the axiomatic system presented in Section 3.2.

Before we state and prove Proposition (2.2), it will be useful for the reader to interpret the statement

$$V_{(i,j),(j,k)}(t) = 1, \quad (i,j) \in A\ ,\ (j,k) \in A$$

to mean that "at time t we have visited node j from node i and will now visit node k from node j."

Proposition (2.2)

Let $u = 1_{\{h\}}$. Each feasible flow to support output u for the production system corresponds to a Hamiltonian cycle in H. Conversely, to each Hamiltonian cycle in H we may associate a feasible flow to support output level u.

Proof of Proposition (2.2)

We first show that a feasible flow corresponds to a Hamiltonian cycle. Assume a feasible flow to support u exists and select one. By (2.21) there exists exactly one $(k,l) \in A$ such that

$$y_{(k,l)}(0) = 1 . \tag{2.25}$$

Since node l must satisfy (2.22), it follows that there exists an m, $1 \leqslant m \leqslant N$, and a time t^* such that

$$V_{(k,l),(l,m)}(t^*) = 1. \tag{2.26}$$

Since no inventories are allowed, this in turn implies that

$$W_{(l,m)}(t^*) = 1. \tag{2.27}$$

Constraint (2.22) implies that

$$\int_{[0,h]} W_{(i,j)} d\mu \leqslant 1, \quad \forall (i,j) \in A. \tag{2.28}$$

Hence, by (2.26) and (2.27) it is easily seen that $\tau(W_{(l,m)}) = t^*$. By definition of $F_{(l,m)}$ (see (2.24)) we have that

$$[F_{(l,m)}(y_{(l,m)}, W_{(l,m)})](t^* + d_{(l,m)}) = 1 \tag{2.29}$$

Since no inventories are allowed, (2.29) implies that

$$\sum_{\{n \mid (m,n) \in A\}} V_{(l,m),(m,n)}(t^* + d_{(l,m)}) + V_{(l,m),(N+1,N+1)}(t^* + d_{(l,m)}) = 1. \tag{2.30}$$

By (2.23), we have that for exactly one n, $1 \leqslant n \leqslant N$,

$$V_{(l,m),(m,n)}(t^* + d_{(l,m)}) = 1 \tag{2.31}$$

which in turn implies that

$$W_{(m,n)}(t^* + d_{(l,m)}) = 1. \tag{2.32}$$

It is clear by the above development that the *forward-through-time argument* continues and traces out a path in G. We symbolize this path by

$$(k,l) \rightarrow (l,m) \rightarrow (m,n) \rightarrow \cdots$$

As an easy consequence of constraint (2.22), the path as generated by our forward-through-time argument must be finite. (Loosely worded, Constraint 2.22 insures that a node in H can not be visited more than once for any feasible flow.) Hence, the path must be of the form

$$(k,l) \rightarrow (l,m) \rightarrow (m,n) \rightarrow \cdots \rightarrow (w,x) \rightarrow (x,y) \rightarrow (y,z).$$

The forward-through-time argument clearly indicates that $F_{(y,z)} \neq 0$. Since no inventory is allowed *except* at the sink node and (y,z) is the last node in the path it is easy to see that $(y,z) = (N+1,N+1)$. By (2.23), it follows that $(x,y) = (k,l)$. Hence, it is easy to verify that

$$W_{(k,l)}(t^* + d_{(l,m)} + d_{(m,n)} + \cdots + d_{(w,k)}) = 1 \tag{2.33}$$

$$F_{(N+1,N+1)}(t^* + d_{(l,m)} + d_{(m,n)} + \cdots + d_{(w,k)}) = 1. \tag{2.34}$$

We see that the forward-through-time argument constructs from the feasible flow a cycle in H which we symbolize by

$$k \rightarrow l \rightarrow m \rightarrow n \rightarrow \cdots \rightarrow w \rightarrow k.$$

We must now argue that this cycle is in fact a Hamiltonian cycle. Intuitively, Constraint 2.22 insures that a node in H has to be visited at least once for any feasible flow.

To argue more formally, let i denote a node in H not in the cycle constructed above. Since i must satisfy Constraint (2.22), it follows that for some h, j and t^{**}

$$V_{(h,i),(i,j)}(t^{**}) = 1. \tag{2.35}$$

Instead of using a forwards-through-time argument starting at t^{**}, we now use a *backwards-through-time arguments.* Constraint (2.35) implies that

$$[F_{(h,i)}(y_{(h,i)}, W_{(h,i)})](t^{**} - d_{(i,j)}) = 1. \tag{2.36}$$

By (2.21), $y_{(h,i)} = 0$, hence by (2.36) $W_{(h,i)}(t^{**} - d_{(i,j)}) = 1$. This means that for some g

$$V_{(g,h),(k,l)}(t^{**} - d_{(i,j)}) = 1. \tag{2.37}$$

It is easy to see how to propagate the backwards-through-time argument.

The path in H generated from the backwards-through-time argument, by (2.22), must terminate. Furthermore, no node in this path can be a node in the cycle we generated from the forward-through-time argument (again, by (2.22)). If (a,b) denotes the first arc in the path generated by the backwards-through-time argument, it is easy to check (and intuitively clear) that

$$y_{(a,b)}(0) = 1. \tag{2.38}$$

And, of course, if $(a,b) \neq (k,l)$ then (2.21) is violated; if $(a,b) = (k,l)$ then (2.22) is violated for node l. Hence, our cycle is Hamiltonian.

Finally, we wish to argue that t^* is in fact equal to $d_{(k,l)}$. *If so, then we will have shown that the intermediate product transfer into the sink node is the indicator of the length of the Hamiltonian cycle.*

By (2.26), we must have that

$$[F_{(k,l)}(y_{(k,l)}, W_{(k,l)})](t^*) = 1. \tag{2.39}$$

Since $W_{(k,l)} = t^* + d_{(l,m)} + d_{(m,n)} + \cdots d_{(w,k)}$ (2.33), by definition of $F_{(k,l)}$ (see 2.24) it is clear that either $t^* = d_{(k,l)}$ or $t^* = t^* + d_{(k,l)} + d_{(l,m)} + \cdots + d_{(w,k)}$. Since $d_{(i,j)} > 0, \forall (i,j) \in A$ we have that $t^* = d_{(k,l)}$.

The conclusion is that a feasible flow corresponds to a Hamiltonian cycle. It is immediate that a Hamiltonian cycle corresponds to a feasible flow. One simply chooses one arc in the cycle as a starting point, say (k,l), and sets $y_{(k,l)}(0) = 1$. If (l,m) is the next arc in the cycle, then one sets

$$[F_{(k,l)}(y_{(k,l)}, W_{(k,l)})](d_{(k,l)}) = 1$$

$$V_{(k,l),(l,m)}(d_{(l,m)}) = 1$$

and continues in the obvious way. That the inventory balance constraints (2.1)-(2.11) are met is immediate. The proof of Proposition (2.2) is now complete.

Let LN refer to the network dynamic production correspondence for the production system described in Proposition (2.2). Then as an immediate corollary we have the following fact.

Corollary (2.3)

For all $\tau \in R_+$, if $u = 1_{[\tau]}$ then $\mathrm{LN}(u) \neq \varnothing$ if and only if there exists a Hamiltonian cycle in H of length less than or equal to τ.

As a consequence of our above developments, the following problem is equivalent to the Traveling Salesman Problem:

$$P:\ \min\left\{\int_{[0,h]} u\,d\mu \mid \exists x \text{ with } x \in \mathrm{LN}(u)\right\}.$$

Hence, this last problem is NP-complete. Since finding the *longest* Hamiltonian cycle is also NP-complete, then the general problem of maximizing *scalar* output for a production system is NP-complete.

2.3.3. Material Requirements Planning

Underlying Assumptions

Material Requirements Planning (MRP) is a production planning tool for the discrete parts manufacturing environment.[2] To use MRP, the production planner assumes that:

(1) All flows, except the intermediate product transfer flows, are step functions associated with an equal length period time grid. The intermediate product transfer flows are assumed to be event-based flows. The number of periods p

[2] For a comprehensive presentation see Orlicky [1970,1975] or Plossl, Wight [1971].

is assumed finite.

(2) *Bill-of-Materials Coefficients* a_{lk}, $l = 1,2,\ldots,m$, $k = 1,2,\ldots,m$ exist such that to produce 1 unit of product k requires a_{lk} units of product l.

(3) Constants L_k, $k = 1,2,\ldots,m$ exist such that it takes no more than L_k periods to produce an arbitrary quantity of product k *regardless* of the magnitude of that quantity.

(4) If a production planner uses MRP, then the planner is tacitly assuming, by (2) and (3), that the production network associated with the discrete parts manufacturing environment is one for which activity A_k represents the production of product k.

Constraints on the Intermediate Product Transfers

If activity A_k is to produce $\int_{\{I_m\}} V^k d\mu$ units of product k in period I_m, then it is required that all intermediate products necessary to produce $\int_{\{I_m\}} V^k d\mu$ units of product k for A_l *must* be available L_k periods earlier. That is, for $1 \leqslant r \leqslant p$, $1 \leqslant j \leqslant m$, $1 \leqslant l \leqslant m$,

$$\sum_{m=1}^{r} a_{lk} \int_{\{I_m\}} V^k d\mu \leqslant \sum_{m=-L_l}^{r} \int_{\{I_{m-L_l}\}} \left\{ \sum_j V^l_{j\mu} \right\} d\mu . \tag{2.40}$$

(2.40) could be suitably modified to incorporate transfer lags if necessary.

Explosion of Requirements

Let u denote a final output vector. If one insists on equality in Constraint (2.40), then it is easy to check that Assumption 4 implies that all V_{ij}'s are determined. This determination is usually referred to as the *material requirement's explosion* or the *explosion of requirements.*

The Fundamental Problem with MRP

The use of MRP only determines a *schedule* for the V_{ij}'s. It says nothing about how the individual activities which comprise the production system will be able to produce the requirements. Stated differently, by Assumption 3, MRP ignores the activity production functions since it assumes that each activity may be able to produce any quantity required. By choosing to ignore the activity production functions, the *production lags*, the L_k's, are introduced and *inflated* to insure feasibility of the schedule of the V_{ij}'s. It is well-known that this is the fundamental problem with MRP.[3]

2.3.4. Single-Project Production Systems

A *single-project production system* is one for which activities are assumed to perform a one-time job, uninterrupted from start to finish, and for a known fixed duration. A *project* is said to be completed when all of the activities are finished. Following Leachman [1983], it is assumed that each exogenous resource is non-storable, that the total amount of each exogenous resource used to complete an activity is known, and that the resources are applied at constant rates.

Associated with a single-project production system is an *activity-on-node precedence network* denoted by H. H is an acyclic, directed graph on N nodes. Node A_i in H corresponds to activity i, $i = 1,2, \ldots, N$. The arcs indicate *strict precedence relationships* between the activities. That is, if there is an arc from A_i then activity j cannot start until activity i has finished. Let W denote the set of whole numbers. Let d_i denote the duration of activity i, $d_i \in W$ for each i.

Definition (2.41)

A *feasible schedule of start-times* for the activities is a vector $S = (S_1, \ldots, S_N) \in W_+^N$ such that if arc (i,j) exists in H then $S_i + d_i \geq S_j$. We note that by taking $S \in W_+^N$, we are assuming that a start-time for an activity always begins at the beginning of a period.

[3] For a detailed description of the problems with MRP, see Kasper [1983].

In this section, a production system is constructed such that, for an appropriate choice for the final output vector u, there is a 1-1 correspondence between the set of feasible flows to support output level u and the set of feasible schedules of start-times. Furthermore, certain feasible flows may be identified with the well-known early-start and late-start schedules determined by the ordinary critical path method.[4]

We first proceed to construct the production network G associated with the production system. To the activity-on-node network H add sink node A_{N+1} and add arcs $(i,N+1)$ from each node A_i to node A_{N+1}. This network will be taken to be the production network G.

To define the production system associated with G, we need to specify the limitations on the primitive elements. We list these below:

(1) The time grid T is the set $\{0,1,2,\cdots\}$.

(2) All flows are step-functions except the final output vector which is event-based.

(3) No disposal of any kind is allowed.

(4) Attention is restricted to the finite horizon $[0,h]$ where $h=\sum_{i=1}^{N} d_i$.

The additional constraints imposed on the flows--other than those constraints already imposed by the general model--are listed below.

On the Applications of System Exogenous Inputs

As notation, let b_{ki}, $k=1,2,\ldots,n$ denote the total amount of resource k that A_i requires. As notation, let, for $i=1,2,\ldots,N$,

$$Z_i=\left\{z_i \in L_+^\infty(\mu) \mid \exists S_i \in R_+ \text{ such that } z_i=\frac{1}{d_i}\cdot 1_{(S_i,S_i+d_i)}\right\} \tag{2.42}$$

and define for $i=N+1$

[4] See Moder and Phillips [1970], for example.

$$Z_{N+1} = \left\{ z_{N+1} \in L_+^\infty(\mu) \mid \exists S_{N+1} \in R_+ \text{ such that } z_{N+1} = 1_{(S_{N+1}, h)} \right\}. \tag{2.43}$$

Then it is required that there exist a $z_i \in Z_i$ such that

$$v_i^k = b_{k_i} z_i, \quad 1 \leq k \leq n. \tag{2.44}$$

Note that the description of the application vector fits the form as required by DLAAM presented in Section 2.3.1. The variable z_i is referred to as an *operating intensity*. The interpretation of z_i is that $\int_0^\tau z_i d\mu$ expresses the fraction of the total amount of each resource required to complete A_i up to time τ or expresses the fraction of the job completed up to time τ.

On the Activity Production Functions

To define $F_i(y_i, W_i)$, one needs to know what kinds of "products" A_i is "producing." It is assumed that each activity produces a *distinct* product for each of the activity's immediate successors. Following DLAAM, the output rate is modeled as the rate of the utilization of the system exogenous inputs. That is,

$$F_i^l(y_i, W_i) = \begin{cases} z_i & \text{if arc } (i,l) \text{ is in } G \text{ and } y_i^k = b_{k_i} z_i \text{ for each k} \\ 0 & \text{otherwise.} \end{cases} \tag{2.45}$$

In effect, (2.45) says that, up to time τ, $x\%$ of an activity has been completed if, up to time τ, $x\%$ of the resources required to complete the activity have been utilized.

On the Intermediate Product Transfers

Each activity A_i, $i \neq N+1$, obtains a distinct intermediate product input from each of its immediate successors.

Definition (2.46)

An *intermediate product transfer functional* is a map $f_{ji}: Z_j \rightarrow Z_i$ defined by

$$f_{ji}\left(\frac{1}{d_j}\cdot 1_{(S_j, S_j+d_j)}\right) = \begin{cases} \frac{1}{d_i}\cdot 1_{(S_j+d_j, S_j+d_j+d_i)} & \text{if } i \neq N+1 \\ 1_{(S_j+d_j, h)} & \text{if } i = N+1 \end{cases} \tag{2.47}$$

It is then required that if

$$y_j^k = b_{kj} z_j\,, \quad 1 \leq k \leq n \tag{2.48}$$

$$y_i^k = b_{ki} z_i\,, \quad 1 \leq k \leq n \tag{2.49}$$

then for $j = 1,2,\ldots,n$, $i = 1,2,\ldots,N+1$,

$$V_{ji} = f_{ji}(z_j) \tag{2.50}$$

$$W_i^j = z_i\,. \tag{2.51}$$

This completes the description of a single-project production system. A routine check verifies that the axioms are satisfied. As notation, let F denote the subset of indices $\{1,2,\ldots,m\}$ associated with the intermediate product transfers into the sink node. The following proposition and its corollaries show why we have chosen to define the primitive elements in the manner given above.

Proposition (2.4)

If the final output vector u is defined by $u^i = 0$ if $i \notin F$ and $u^i = 1_{\{h\}}$ if $i \in F$, then there is a 1-1 correspondence between the set of feasible flows to support u and the set of feasible schedules of start-times.

Proof of Proposition (2.4)

Assume feasible flows to support output u exist and select one. Each y_i, by constraint (2.44), has an operating intensity associated with it. By definition (2.42), each z_i has an

$S_i \in R_+$ associated with it. We claim that the vector $S = (S_1, \ldots, S_N)$ so generated is a feasible schedule of start-times.

Suppose A_j is an immediate predecessor to A_i. By the definition of a feasible flow the following inventory balance constraint must hold:

$$\int_0^\tau \{V_{ji} - W_i^j\} d\mu \geq 0, \quad \forall i, \forall \tau \in [0,h]. \tag{2.52}$$

By the constraints imposed on V_{ji} and W_i^j (see (2.50), (2.51)) it follows that upon substitution

$$\int_0^\tau \left\{\frac{1}{d_j} 1_{(S_j+d_j, S_j+d_j+d_i)} - \frac{1}{d_i} 1_{(S_i, S_i+d_i)}\right\} d\mu \geq 0, \quad \forall \tau \in [0,h]. \tag{2.53}$$

Clearly this implies that

$$S_i \geq S_j + d_j. \tag{2.54}$$

Hence, S is a feasible schedule.

Conversely, if S is a feasible schedule of start-times then define the *induced operating intensity for A_i*, denoted by z_i, to be

$$z_i^S = 1_{(S_i, S_i+d_i)}. \tag{2.55}$$

A routine check verifies that the primitive elements defined from the operating intensities given in (2.55) satisfy the inventory balance constraints of the general model (2.1-2.11). Thus, a feasible schedule S induces a feasible flow in a natural way.

The association between feasible flows and feasible schedules given above is easily seen to be 1-1. The proof is now complete.■

Definition (2.56)

An *Early-start schedule*, denoted by E, is a feasible schedule such that if S is any other feasible schedule then $S_i \geqslant E_i$ for $i = 1,2, \ldots, N$.

Definition (2.57)

A *Late-start schedule*, denoted by L, is a feasible schedule such that if S is any other feasible schedule then $S_i \leqslant L_i$ for $i = 1,2, \ldots, N$.

The following two corollaries are immediate:

Corollary (2.5)

A feasible flow minimizes the sum, over all activities, of the cumulative intermediate product inventories at the sink node if and only if the schedule of start-times associated with feasible flow is the late-start schedule.

Corollary (2.6)

A feasible flow maximizes the sum, over all activities, of the cumulative intermediate product inventories at the sink node if and only if the schedule of start-times associated with the feasible flow is the early-start schedule. For ease of presentation, we excluded it.

2.4. Comments on the General Model

Comments about the general model and extensions to the model are made below.

On a Stochastic Framework

By our assumption and definition of the activity production functions, our framework is clearly deterministic. We are assuming that if we *knew* the applications of input that we would *know* the realized output obtained from production. Environments such as agriculture where weather plays an important role in determining yield do not fit this assumption.

One can make the production functions *random functionals* to account for such environments. If we let Ω denote some probability space[1] then we can define F_i as a map from

$$L_{y_i} \times L_{W_i} \times \Omega \rightarrow R_+$$

where one interprets $F_i(y_i, W_i, w)$ to be the actual output if (y_i, W_i) was the vector of inputs applied and w was the state of nature observed. We are tacitly assuming here that Ω is *independent* of $L_{y_i} \times L_{W_i}$, i.e., Ω does not vary with the choice of (y_i, W_i). Axioms on F_i would have to be suitably modified to account for the stochastic formulation of F_i.[2] To model production systems where it is more appropriate to view Ω as dependent on the choice of (y_i, W_i) is considerably more difficult but worth investigating.

Lags

As we have already mentioned, no explicit mention of time lags was given in the general model. Such lags could be easily incorporated but would require the introduction of new variables in a manner similar to that treated in Section 2.1.4. There, variables V_{ij}^* were introduced.

---

[1] $\Omega = (Y, B, P)$ where Y is a set, B is a σ-algebra of subsets of Y and P is a measure such that P is σ-finite with $P(Y) = 1$.

[2] Mak [1981] develops a stochastic theory of Dynamic Production Correspondences.

Dependence Between Activity Primitive Elements

As illustrated in our description of the Traveling Salesman Problem, specific production systems might have additional constraints linking primitive elements belonging to different activities. By changing our definition of the domains, we could allow for this flexibility. For ease of presentation, we excluded it.

3. TOPICS IN PRODUCTION THEORY

This chapter addresses some of the issues found in Production Theory. In Section 3.1 we analyse the *technically efficient subset.* In Section 3.2 we carefully analyse Shephard's duality in both the steady-state and dynamic cases. Section 3.3 addresses 2 versions of the Law of Diminishing Returns as first formulated by Shephard. Finally, in the Appendix, Section 3.4, some of the technical propositions are proved.

3.1. On The Technically Efficient Subset

3.1.1. On the Nonemptiness of the Technically Efficient Subset

The *technically efficient subset,* denoted by $\mathrm{EN}(u)$, is defined to be

$$\mathrm{EN}(u) = \left\{ x \in \mathrm{LN}(u) \mid \text{if } y \leqslant x \text{ then } y \notin \mathrm{LN}(u) \right\}.^{1} \tag{3.1}$$

The technically efficient points are those for which it is not possible to lower the input rates and still achieve the same output rates. Thus it is important to know that our general framework is consistent with a desirable property that the efficient subset is nonempty whenever the level set itself is nonempty. This result is also the first stepping stone towards proving Shephard's Duality Theorem in the finite horizon dynamic case which is presented in the next section.

We remark that we prove this theorem under the assumption of a *finite* horizon.[2] That is, for some $h > 0$ the underlying space of flows of goods and services as presented in Axiom 1, Chapter 3 will be changed to $L_+^\infty([0,h], B, \lambda + \nu)$ where B is the restriction of the Lebesgue *σ−field* to $[0,h]$, λ is Lebesgue measure restricted to $([0,h], B)$ and ν the counting measure on $\{t_k \mid t_k \leqslant h\}$. As notation, let $L_+^\infty(\mu)$ denote $L_+^\infty([0,h], B, \lambda + \nu)$.

---

[1] If $y \leqslant x$, then $\forall i$, $1 \leqslant i \leqslant n$, $\mu\{x^i > y^i\} = 0$.

[2] We do not have a proof in the infinite horizon case. Additional hypotheses on the primitive elements we believe would have to be taken.

Theorem 3.1.1

For the finite horizon case, if $\mathrm{LN}(u) \neq \varnothing$ then $\mathrm{EN}(u) \neq \varnothing$.

Proof of Theorem (3.1.1)

Pick an arbitrary z in the nonempty $LN(u)$. Consider the following optimization problem.

$$P(z):\quad \inf \int \{\sum_j x^j\} d\mu$$

subject to:

(1) $x \in \mathrm{LN}(u)$

(2) $x \leqslant z$

It is clear that if we can show that there exists an $x \in \mathrm{LN}(u)$, $x \leqslant z$ achieving the infimum in $P(z)$ then $x \in \mathrm{EN}(u)$ and $\mathrm{EN}(u)$ is nonempty. What we have to show is that this problem has a solution.

To show this, endow each L_i (the subset of $L_+^\infty(\mu)$ satisfying Axioms 2-4 for the i^{th} flow type) with the relative weak-star topology (viewing it as a subset of $L^\infty(\mu)$) and $L_x^1 \times \cdots \times L_x^n$ with the product topology. Letting L_x stand for $L_x^1 \times \cdots \times L_x^n$ then the objective function ϕ viewed as a map from $L_x \xrightarrow{\phi} R_+$ is continuous.[3] As an application of the theorems of Banach-Alaoglu and Tychonoff, the set of x in L_x satisfying (2) is contained in a compact subset of $(L^\infty(\mu))^n$.[4] In the Appendix we verify that each L_{x_i} is closed in $L^\infty(\mu)$ (Proposition 3.4.2); hence, the set of x in L_x satisfying (2) is compact in L_x. Since a continuous real-valued function on compact set achieves its infimum then in order to complete the proof that $\mathrm{EN}(u)$ is nonempty it suffices to show that $\mathrm{LN}(u)$ is a weak-star closed subset of $(L^\infty(\mu))^n$.[5]

---

[3] The operation of addition is continuous. The integrable function here is $p = 1_{[0,T]}$.

[4] The theorem of Banach-Alaoglu states that the closed unit ball in the dual space of an arbitrary topological vector space is weak-star compact (Rudin [1973], p. 66). Since L^∞ is the dual to L^1 we have that $\{x \mid \|x\|_\infty \leqslant 1\}$ is weak-star compact. The theorem of Tychonoff is that an arbitrary product of compact spaces is compact in the product topology (Dugundgi [1966], p. 224).

[5] See Dugundgi, p. 227.

To show that $\mathrm{LN}(u)$ is closed let $\{x_\alpha\}$ be a net in $\mathrm{LN}(u)$ converging to x.[6] To say that $x_\alpha \in \mathrm{LN}(u)$ for all α means that for all α there exists a feasible flow for x_α to support output level u. In order to show that x is in $\mathrm{LN}(u)$ (which is our goal) we need to exhibit a feasible flow for x to support output level u. To arrive at this end we will extract "limit" functions for each flow type from the feasible flows for each x_α.

By the very definition of a feasible flow the net $\{x_\alpha\}$ induces a net of functions for each flow type. What we will in effect show is that there exists a subnet of the original net $\{x_\alpha\}$ such that the associated sub-nets of functions for each flow type have the property that they are uniformly bounded in the $L^\infty(\mu)$ norm. Hence, by compactness, we may extract these sub-nets to obtain limit functions.[7] By duality between $L^1(\mu)$ and $L^\infty(\mu)$, the inventory balance constraints will hold for these functions.[8] The proof would then be complete.

The inventory balance constraints which must be satisfied by the various pairs of flow types which when taken together make a flow feasible have the following structural form:

$$0 \leq S + \int_0^\tau \{x - T - y\} d\mu \leq C, \quad \forall \tau \in [0,h]. \tag{3.2}$$

Here S represents the initial stock, if any, C the capacity (perhaps infinite), x the flow of input into the "system", T the flow of disposal out of the system, and y the flow of input out of the system. (See Section 3.2.3.)

To arrive at our main result we first prove a simple proposition which paves the way for a "verbal" proof of the rest of the theorem.

---

[6] We use nets instead of sequences since $L^\infty(\mu)$ as topologized does not have a countable neighborhood base. For definition of a net see Halmos [1974], p. 65. For those unfamiliar with nets think of sequences instead.

[7] We know that the closed unit ball is compact. Operations of scaling are homeomorphisms in any topological vector space.

[8] We are using the fact that $L^1(\mu)$ is the pre-dual of $L^\infty(\mu)$ hence with $p = 1_{[0,h]}$ the convergence is maintained.

Proposition (3.1.2)

Let $\{x_\alpha\}$, $\{y_\alpha\}$, and $\{T_\alpha\}$ be nets in the (appropriate) spaces L_x, L_y, and L_T which for each α satisfy (3-2). Further assume that $\{x_\alpha\} \rightharpoonup x$. Then subnets of the original nets exist which are *uniformly* bounded in the $L^\infty(\mu)$ norm.

Proof of Proposition (3.1.2)

Let $\epsilon > 0$. For some β we have that eventually $\forall\alpha \geqslant \beta$, $\int x_\alpha d\mu \leqslant \int x d\mu + \epsilon$ by convergence. Thus, eventually, $\int y_\alpha d\mu \leqslant \int x d\mu + \epsilon$ and $\int T_\alpha d\mu \leqslant \int x + \epsilon$. By Axiom 2 it now follows easily that eventually the norms of y_α and T_α (and of course x_α) must be *uniformly* bounded. ■

Now on to the conclusion of the proof of the theorem. First treat the storage case. Since $x_\alpha^j \rightharpoonup x^j$ for each j it follows easily by Proposition (3.1.2) that we may find an appropriate subnet so that the x_{0i}^j and y_i^j sub-nets are uniformly bounded. By Axiom 8, each of the induced $F_i^*(y_i^\alpha, W_i^\alpha)$ sub-nets are uniformly bounded too. Since there is an inventory balance constraint of type (3.2) linking the $F_i^*(y_i^\alpha, W_i^\alpha)$*'s* with the $\sum_{j=1}^{j=N+1} V_{ij}^\alpha$*'s* it follows by Proposition (3.1.2) that a further refinement of the original net gives us a uniform bound on the V_{ij} nets.[9] The same argument may be used to deduce the same result for the W_i and T_i sub-nets. Now since all nets of functions of each flow type are contained in compact subsets we may extract *convergent* sub-nets.

If y_i and W_i denote, for each i, the limit vectors of the sub-nets $\{y_\alpha\}$, $\{W_i^\alpha\}$, then by Proposition (3.4.6) in the Appendix to this Chapter we see that

$$F_i^*(y_i^\alpha, W_i^\alpha) \rightharpoonup F_i^*(y_i, W_i). \tag{3.3}$$

[9] Actually the same result could be obtained if we invoked Axiom 12. This would immediately imply that the nets $\{F_i^*(y_i^\alpha, W_i^\alpha)\}$, $\{V_{ij}^\alpha\}$, and $\{W_i^\alpha\}$ are uniformly bounded (since we are restricted to a finite horizon). See the comments at the end of Axiom 12.

In view of (3.3) it now follows easily that the limit functions together satisfy all of the inventory balance constraints needed to define a feasible flow for x to support output level u. Hence, $x \in \mathrm{LN}(u)$ which implies that $\mathrm{LN}(u)$ is closed as desired in the storage case.

If inventories (in some cases) were not allowed, then the argument is the same except that it is much easier to obtain uniform bounds for the sub-nets. One needs to insure however that if $\{x_\alpha\}$, $\{y_\alpha\}$ are nets such that $x_\alpha \leqslant y_\alpha$ with $x_\alpha \to x$, $y_\alpha \to y$ for all α then x is indeed less than or equal to y. This simple verification may be found in the Appendix, Proposition (3.4.1). Our proof to Theorem 3.1.1 is now complete.■

An immediate corollary, which will be needed later, is given below.

Corollary to Theorem (3.1.1)

$$\mathrm{LN}(u) \subset \mathrm{EN}(u) + (L_+^\infty(\mu))^n \tag{3.4}$$

We remark that we used $(L_+^\infty(\mu))^n$ instead of L_x since the difference of two functions in L_x need *not* be in L_x.

3.1.2. Extensions to Theorem (3.1.1)

If x^* were a solution to problem $P(z)$, then x^* would be a vector of system exogenous inputs which minimizes the total amount of resource over all x in $\mathrm{LN}(u) \cap \{x \mid x \leqslant z\}$. It is certainly possible that x^* might *not* minimize the total amount of resource over all x in $\mathrm{LN}(u)$. If $\phi\colon L_x \to R_+$ denotes the objective function, then it would be desirable to show that the problem defined by

$$\begin{aligned} \mathrm{P}: \quad & \inf \phi(x) \\ & \text{subject to:} \quad x \in \mathrm{LN}(u) \end{aligned}$$

has a solution.

Actually, this is an easy consequence of Axiom 2 as we now show. Suppose one could find an x which was a solution to P. Let $\{g_k^j\}$, $\{h_k^j\}$ denote the set of g_k's and h_k's (as defined in Axiom 2) associated with the flow type $\{x^j\}$. Since x is assumed to be a solution to P and in view of Axiom 2 it is easily seen that

$$\|x^j\|_\infty \leq \max_{\{k \mid t_k \leq T\}} \{g_k^j(\phi(z)), h_k^j(\phi(z))\}$$

$$\equiv f(\phi(z)).$$

Hence, $\|x\|_\infty = \max_j \|x^j\|_\infty \leq f(\phi(z))$. If we were to define the following problem

P^*: $\inf \phi(x)$

subject to:

(1) $x \in \mathrm{LN}(u)$

(2) $\|x\|_\infty \leq f(\phi(z))$,

then by the above it may be seen that any solution to P^* is a solution to P. But P^* has at least *one* solution; hence, so does P.[10]

We may say more about problem P if we note that the objective function ϕ in P has the following structural form:

$$\int \sum_{j=1}^{j=n} (p^j \cdot x^j) d\mu \tag{3.5}$$

where the p^j's were all taken to be the function $1_{[0,h]}$. For any $p = (p^1, \ldots, p^n) \in (L^1(\mu))^n$ the functional expression given in (3.5) viewed as a map from $L_\infty \to R$ is continuous (by duality). Hence, a solution to P for general $p \in (L^1(\mu))^n$ *exists*. If $p \in (L^1(\mu))^n$ was strictly positive, i.e., $p^j > 0$ for each j, then any point achieving the infimum would clearly be *efficient*.

[10] Use the proof of Theorem (4.1.1). We only used the fact that since $x \leq z$ then we could restrict $\mathrm{LN}(u)$ to a compact subset. We get the same result by directly restricting the set $\mathrm{LN}(u)$ to a closed ball.

Let $(L^1(\mu))^n_{++}$ denote the subset of $(L^1(\mu))^n$ of strictly positive vectors and let $L_u = L_{u^1} \times \cdots \times L_{u^m}$. Let us define a correspondence from $(p,u) \in (L^1(\mu))^n_{++} \times L_u \to \psi(p,u) \in L_x$ where one defines $\psi(p,u)$ as the set of all minimizers to Problem P with the objective function as in (3-5). The p^j's may be thought of as weighting factors for the resources which take into consideration their relative value to the production planner and the fact that future costs may need to be discounted.[11] By our above comments we know that each point in $\bigcup_{\{p \in (L^1(\mu))^n_{++}\}} \psi(p,u)$ is efficient for LN(u) and has the property that it can be obtained through minimization of an appropriate optimization problem. Alternatively, and on more geometric grounds, one can say that each point in $\bigcup_{\{p \in (L^1(\mu))^n_{++}\}} \psi(p,u)$ has the property that there exists a *hyperplane* which supports LN(u) at this point.[12] An immediate theoretical question is whether all efficient points can be obtained in this way, i.e., is

$$\bigcup_{\{p \in (L^1(\mu))^n_{++}\}} \psi(p,u) = \mathrm{EN}(u)\,?$$

If this were true, then we would have a convenient characterization of the efficient subset.

A simpler theoretical question to address is whether or not

$$\bigcup_{\{p \in (L^1(\mu))^n\}} \psi(p,u) \supset \mathrm{EN}(u),$$

i.e., does each efficient point have a hyperplane which supports LN(u) at the point (with p not necessarily strictly positive)? In the next section, we construct an example which shows that in the *infinite* horizon case the answer to this question is no-- even if we assume that the level set is *convex* and *compact*. The counter-example, being constructed for the infinite horizon case, shows that even if additional hypotheses were imposed on the primitive elements so that

[11] In fact future costs were *not* discounted in the previous example. Further, all of the resources were given equal "weights".

[12] If X is a Topological Vector Space, then a hyperplane associated with X is a subset of X generated by the following equation: $\{x \mid f(x) = c\}$ where $f \in X^*$ (the set of continuous linear functionals on X) and $c \in R$.

Theorem (3.1.1) could be extended to the infinite horizon case efficient points need not be obtainable through an optimization problem of type (3.5). We remark that the level sets would be convex if the production functions were assumed to be concave. In the Dynamic Linear Activity Analysis Model of Production (DLAAM) discussed in Section (2.3.1), the production functions were concave.

3.1.3. Counter-Example: Efficient Points Need Not Be Supportable[13]

The mathematical statement to the question raised in the previous section may be phrased as follows. Let $S \subset (L_+^\infty(\mu))^n$ be a (weak-star) compact, convex, *monotonic* subset. By monotonicity we mean that if $x \in S$ and $y \geqslant x$ then $y \in S$. Let x be an efficient point of S, denoted by $x \in E(S)$. Does there exist a $p \in (L^1(\mu))^n$ such that

$$\sum_{j=1}^{j=n} \int p^j \cdot x^j \, d\mu \leqslant \sum_{j=1}^{j=n} \int p^j \cdot z^j \, d\mu, \quad \forall z \in S?$$

As mentioned in the previous section, the counter-example is constructed in the infinite horizon case (that is, we now work with the measurable space $(R_+, \mathcal{B})$ instead of $([0,h], \mathcal{B})$. To construct the counter-example we will first work in the space l^∞ and then "translate" into $L_+^\infty(\mu)$. (It suffices to construct an example in the case when $n = 1$).

Define ϵ_n in l^∞ by

$$\epsilon_n^j = \begin{cases} 1 & \text{if } j = n \\ -1 & \text{if } j = n-1 \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

Let $S = co\{\epsilon_n\}_1^\infty$ denote the convex hull of the ϵ_n's.[14] Since the closure of a convex set in a topological vector space is also convex then $\bar{S}$ is closed and convex.[15] We will show below that

[13] I wish to gratefully acknowledge the assistance of Stephen Boyd, Department of Electrical Engineering and Computer Science, University of California at Berkeley, in the construction of the counter-example.

[14] If $z \in S$, then there exists a $\lambda \in l_+^\infty$, $\|\lambda\|_1 = 1$ with the number of positive indices finite such that $z = \sum \lambda^k \epsilon_k$, i.e., for each j, $z_j = \sum \lambda^k \epsilon_k^j$.

[15] Rudin [1973], p 11

$\bar{S}$ is bounded in norm by 1; it would then follow that $\bar{S}$ is compact. We will show that $0 \in E(\bar{S})$ but 0 is *not* supportable in the sense described in the previous section.[16]

Proposition (3.1.3)

$0 \in E(\bar{S})$.

Proof of Proposition (3.1.3)

Suppose this were not true. Then this would mean that we could find a $z \in \bar{S}$ such that $z_k \leq 0$ for all k and for some i, $z_i < 0$. We can find a net $\{x_\alpha\}$ in S such that $x_\alpha \to z$. By the definition of convex hull we have that, for some $\lambda_\alpha \in l_+^\infty$ with l^1 norm equal to 1 whose number of positive indices are finite, $x_\alpha = \sum_k \lambda_\alpha^k e_k$ for each α. By the definition of e_n and the fact that $x_\alpha \to z$ it follows that $|(\lambda_\alpha^i - \lambda_\alpha^{i+1}) - z^i| \to 0$ eventually. (This immediately implies that $|z^i| \leq 1$; thus $\|z\|_\infty \leq 1$ which shows now that $\bar{S}$ is bounded in norm by 1.)

The nets $\{\lambda_\alpha^i\}$ and $\{\lambda_\alpha^{i+1}\}$ are each contained in the unit interval which is compact. Hence, each net has a convergent sub-net. Without changing notation for the sub-net let c^i and c^{i+1} denote the limiting values. Clearly, $(c^i - c^{i+1}) = z^i$. Since $z^i < 0$ this implies that $c^{i+1} \geq |z^i| > 0$. Find integer N so that $N|z^i| > 1$.

Now by *sequentially* repeating this process for the nets $\{\lambda_\alpha^{i+r}\}$, $r = 1,2, \ldots, N$ and not changing notation for sub-nets we have that there exist constants c^{i+r}, $r = 1,2, \ldots, N$ satisfying

$$\lambda_\alpha^{i+r} \to c^{i+r}, \quad r = 1,2, \ldots, N \tag{3.7}$$

$$c^{i+r} - c^{i+r+1} = z^{i+r}, \quad r = 1,2, \ldots, N \tag{3.8}$$

Since $z \leq 0$ the constants c^{i+r}, $r = 1,2, \ldots, N$ are non-decreasing. Since $c^{i+r} \geq |z^i|$ this in

[16] By Krein-Millman's Theorem (see Rudin, [1973], p. 70), a closed, convex compact subset of a topological vector space has to be the closure of the convex hull of its extreme points. Here, the e_n's are the extreme points of $\bar{S}$.

turn implies that

$$\sum_{r=1}^{r=N} c^{t+r} \geq Nc^{t+1} \geq N|z^t| > 1 \tag{3.9}$$

Since $\sum_{r=1}^{r=N} \lambda_a^{t+r} \to \sum_{r=1}^{r=N} c^{t+r}$ (by (3.7)) eventually (in view of (3.9)) $\sum_{r=1}^{r=N} \lambda_a^{t+r} > 1$. But then this means that eventually $x_a \notin S$, an obvious contradiction. The result follows.■

Proposition (3.1.4)

0 is not supportable, i.e., there does not exist a $p \in l^1$, $p \neq 0$ such that $0 = p \cdot 0 \leq p \cdot x$, $\forall$ $x \in \bar{S}$.

Proof of Proposition (3.1.4)

Suppose such a p exists. Then $p \cdot e_1 \geq 0$ which implies that $p^1 \geq 0$. Since $p \cdot e_n \geq 0$ for each n this in turn implies that $p^{n+1} \geq p^n$ for each n. Thus the sequence of p^n's is nondecreasing and nonnegative. Therefore, either $p = 0$ or $p \notin l^1$. In either case we would have a contradiction. The result follows.■

We now turn to constructing our counter-example in $L_+^\infty(\mu)$. Let

$$1 + \bar{S} = \{y \in l^\infty \mid \exists x \in \bar{S} \text{ such that } y_i = x_i + 1, \forall i\}. \tag{3.10}$$

Since translations in topological vector spaces are homeomorphisms which preserve convexity and efficiency then 1 is an efficient point of $1 + \bar{S}$ and is *not* supportable in the sense of Proposition (3.1.4). Now we may identify $1 + \bar{S}$ with a compact subset of $L_+^\infty(\mu)$ in the obvious way by associating each $x \in 1 + \bar{S}$ with $x(t) \in L_+^\infty(\mu)$ defined by

$$x = \sum_{i=1}^{\infty} x_i \cdot 1_{(i-1,i)} \tag{3.11}$$

(Here, the set of time points defining the possible times of event-based transfers is the set of nonnegative integers.) Let S^* denote the identification of $1 + \bar{S}$ in $L_+^\infty(\mu)$. Then S^* has the

desired properties we were seeking in our construction of a counter-example.

This example illustrates the problem of additional flexibility available by *infinite* time substitution. The step functions ϵ_i that generate S^* are those which are 1 everywhere except in two consecutive periods $(i-1,i)$, $(i,i+1)$ where the function is 0 on $(i-1,i)$ and 2 on $(i,i+1)$. Interpret the functions in $L^1(\mu)$ as price functions and the optimization problem given in (3.5) as cost minimization. For any price function $p \in L^1(\mu)$, $p \neq 0$, the cost of the function $1 = \sum_{i=1}^{\infty} 1_{(i-1,i)}$ is $\|p\|_1$. Furthermore for some i, $\int p \cdot 1_{(i-1,i)} > \int p \cdot 1_{(i,i+1)}$ (if not, then the proof of Proposition (3.1.4) shows that either $p = 0$ or $p \notin L^1(\mu)$). Hence

$$\begin{aligned}
\textit{Cost of } \sum_{i=1}^{\infty} 1_{(i-1,i)} &= \sum_{k=1}^{\infty} \int p \cdot 1_{(k-1,k)} d\mu \\
&= \sum_{k \neq i,i+1} \int p \cdot 1_{(k-1,k)} d\mu + \int p \cdot 1_{(i-1,i)} d\mu + \int p \cdot 1_{(i,i+1)} d\mu \\
&> \sum_{k \neq i,i+1} \int p \cdot 1(k-1,k) d\mu + 2 \int p \cdot 1_{(i,i+1)} d\mu \\
&= \textit{Cost of } \epsilon_i .
\end{aligned}$$

Thus, ϵ_i would be cheaper than 1 so that 1 could never be obtained from cost minimization, i.e., 1 is never supportable. The reason such a cost tradeoff as illustrated above can always be done is that the generators ϵ_i of S^* "go out towards infinity."

Next, we investigate a property of our counter-example which is of independent interest. Consider the following theoretical question: is it always true that one can separate two compact, convex subsets of R_+^2 which meet at only one point? Answer: No; counter-example: two line segments meeting at a single point. This example may be constructed because we did not insist that the two sets in question have non-empty interior. It turns out that our example also has this property.

Proposition (3.1.5)

$\bar{S}$ is nowhere dense in l^∞, i.e., the interior of $\bar{S} = \varnothing$.

Proof of Proposition (3.1.5)

Let $x \in \bar{S}$. Let $\{x_\alpha\}$ be a net in S converging to x. Write $x_\alpha = \sum_k \lambda_k^\alpha e_k$ where for each α, $\lambda^\alpha \in l^1$, $\|\lambda^\alpha\|_1 = 1$, and the number of positive indices finite. Since $x_\alpha \to x$ this means that $\lambda_1^\alpha - \lambda_2^\alpha \to x_1$. Extract a sub-net of the net $\{x_\alpha\}$ so that the sub-net $\{\lambda_\alpha^1\}$ converges, say to ρ_1. Since $\lambda_1^\alpha - \lambda_2^\alpha \to x_1$ it follows that $\lambda_2^\alpha \to \rho_1 - x_1$. In fact, the following simple claim may be made.

Claim

$$\lambda_k^\alpha \to \rho_1 - \sum_{i=1}^{k-1} x_i , \quad \forall k .$$

Proof of Claim

We have shown this to be true when $k = 1$ and $k = 2$. Suppose it is true for all $k \leqslant m$. Since $\lambda_m^\alpha - \lambda_{m+1}^\alpha \to x_m$ and by induction $\lambda_m^\alpha \to \rho_1 - \sum_{i=1}^{m-1} x_i$ it follows that $\lambda_{m+1}^\alpha \to \lambda_m^\alpha - x_m = \rho_1 - \sum_{i=1}^{m} x_i$. The induction step is shown and the result follows.∎

For all α and for each N we have that $1 \geqslant \sum_{k=1}^{N} \lambda_k^\alpha$ which implies by our claim that for all N, $1 \geqslant \sum_{k=1}^{N} (\rho_1 - \sum_{i=1}^{k-1} x_i)$ and hence $1 \geqslant \sum_{k=1}^{\infty} (\rho_1 - \sum_{i=1}^{k-1} x_i)$. So if we define $\rho_k = \rho_1 - \sum_{i=1}^{k-1} x_i$ then it may be easily seen that $\rho \in l_+^\infty$, $\|\rho\|_1 \leqslant 1$ and the expression $x = \sum_k \rho_k e_k$ is well-defined and correct.

We have thus shown by our above developments that

$$\bar{S} \subset \left\{ x \mid \exists \rho \in l_+^\infty , \|\rho\|_1 \leqslant 1, x = \sum_k \rho_k e_k \right\}.$$

The converse conclusion is simple to show; hence, we may conclude that

$$\bar{S} = \left\{ x \mid \exists \rho \in l_+^{\infty},\ \|\rho\|_1 \leqslant 1,\ x = \sum_k \rho_k \epsilon_k \right\}. \tag{3.12}$$

With this we will now turn to proving the proposition.

If the interior of $\bar{S} \neq \varnothing$, then we can find an $x \in \bar{S}$ and a neighborhood N_x about x such that $N_x \subset \bar{S}$. Find a net in S, $\{x^\alpha\}$, such that $x^\alpha \to x$. Since the weak-star topology is weaker than the norm topology then for some $\epsilon > 0$ we have that eventually, for some α,

$$x^\alpha \in B_\epsilon(x^\alpha) \subset N_x \subset \bar{S}.$$

Clearly this means there is some $x \in \bar{S}$ such that $z \underset{\neq}{\leq} x^\alpha$ (we fix some α large enough) and $z \neq 0$.

If $z \in \bar{S}$, then by (4.12) we have that for some $\rho \in l_+^1$, $\|\rho\|_1 \leqslant 1$ (and number of positive indices finite) $z = \sum_k \rho_k \epsilon_k$, i.e., $z_i = \rho_i - \rho_{i+1}$. Since $x^\alpha \in S$ eventually $x_i^\alpha = 0$. So, eventually, $\rho_i \leqslant \rho_{i+1}$ since $z \underset{\neq}{\leq} x^\alpha$. Since $z \neq 0$, $\|\rho\|_1 \leqslant 1$ we have that eventually $\rho_i \leqslant 0$. So the conclusion is that the number of positive indices of ρ is finite and non-empty (since $z \neq 0$). It is not hard to see that if ρ has the above properties then $\|\rho\|_1 = 1$ so that $z \in S$.

Write x^α as $\sum \gamma_k \epsilon_k$ for some appropriate γ. Since $x^\alpha \underset{\neq}{\geq} z$ we have

$$\gamma_i - \gamma_{i+1} \geqslant \lambda_i - \lambda_{i+1}, \quad \forall i, \tag{3.13}$$

or that

$$\gamma_i - \lambda_i \geqslant \gamma_{i+1} - \lambda_{i+1}, \quad \forall i. \tag{3.14}$$

Since the number of positive indices for *both* γ and ρ is finite eventually $\gamma_{i+1} - \lambda_{i+1} = 0$ so that $\gamma_i \geqslant \lambda_i$, $\forall i$ in view of (3.14). This in turn implies that since $\sum_i \gamma_i = \sum_i \lambda_i = 1$ (due to the finiteness of the number of positive indices) that $\gamma_i = \lambda_i$, $\forall i$. Hence, $x_\alpha = z$ which is a

finiteness of the number of positive indices) that $\gamma_i = \lambda_i, \forall i$. Hence, $x_\alpha = z$ which is a contradiction. The result follows.■

We make one remark. What we have shown is that $S \subset E(\bar{S})$; that is, all points in S are efficient. Also, by (3-12) it is easy to see that not all points in $\bar{S}$ are efficient. For example, take $\gamma_i = \frac{1}{2}^i$, $\lambda_i = \frac{1}{2}(\frac{1}{2})^i$ and let $x = \sum \gamma_k \epsilon_k$, $y = \sum \lambda_k \epsilon_k$. Then x is not efficient.

3.1.4. Compactness of the Closure of the Efficient Subset

To prove Shephard's duality, it is necessary that the closure of the efficient subset be compact. In the axiomatic framework of the Dynamic Theory of Production Correspondences as developed by Shephard and Fare (1980), this property of compactness of the closure of the efficient subset was *assumed as an axiom.* We however prove this property from our axiomatic framework as presented in Section 2.2.

Proposition (3.1.6)

$\overline{\mathrm{EN}(u)}$ is compact.

Proof of Proposition (3.1.6)

Assume that $\overline{\mathrm{EN}(u)}$ is nonempty otherwise there is nothing to show. Let $x \in \mathrm{EN}(u)$ and let $\{x^\alpha\}$ be a net in $\mathrm{EN}(u)$ converging to x. Associate to each x^α a feasible flow for x^α to support output level u. From the proof of Theorem (3.1.1) we know that we may extract a sub-net of the original net $\{x^\alpha\}$ so that the sub-nets associated with each of the flow type nets induced from the feasible flows for x^α to support output level u converge and are uniformly bounded. We denote the sub-net again by $\{x^\alpha\}$.

Since $x^\alpha \in \mathrm{EN}(u)$ this means that there could not be any disposal of system exogenous input in the feasible flow for x^α to support output level u. Thus, for each j we must have that for each α

$$\int_0^h (x^j)^\alpha d\mu = \int_0^h [\sum_i (x^j_{0i})^\alpha] d\mu = \int_0^h [\sum_i (x^j_i)^\alpha] d\mu . \tag{3.15}$$

Furthermore in view of Axiom 10 it follows easily that

$$\int_0^h \{\sum_i (y_i^j)^\alpha d\mu\} < \sum_i g_i(W_i^\alpha, h) \tag{3.16}$$

(else $x^\alpha \notin \mathrm{EN}(u)$). By the assumed properties on g_i and that the net $\{W_i^\alpha\}$ is uniformly bounded, it follows that, for some known function H_i of $B(T)$ and T,[17]

$$\begin{aligned} \int_0^h \{\sum_i y_i^j\} d\mu &\leqslant \sum_i g_i(W_i, h) \\ &\leqslant \sum_i H_i(B(T), T) \quad \text{(by Axiom 12)} \end{aligned} \tag{3.17}$$

where y_i and W_i represent the limit vectors of the nets $\{y_i^\alpha\}$, $\{W_i^\alpha\}$.

Hence by (3-15) it is easy to see that

$$\int_0^h x^j d\mu \leqslant \sum_i H_i(W_i, h)$$

which by Axiom 2 implies that $\|x^j\|_\infty$ is bounded by a known function of $\sum_i H_i(W_i, h)$ which is independent of x^j. Hence, $\|x\|_\infty$ is bounded by a known function of $\sum_i H_i(W_i, h)$ which implies that $\overline{\mathrm{EN}(u)}$ is compact.■

---

[17] See the comments at the end of the statement of Axiom 12, Section 2.2.8. It discusses why in the finite horizon net $\{W_i^\alpha\}$ is uniformly bounded under Axiom 12.

3.2. On Shephard's Duality

3.2.1. Statement of Shephard's Duality Theorem (Steady-State)

In the Shephard model (Shephard, [1970]), the production technology is characterized by the *Production Input sets* or level sets $L(u)$, $u \in R_+$. $L(u)$ is to be interpreted as the set of all input vectors $x \in R^n_+$ which yield at least output rate u. The level sets are assumed to satisfy the following axioms:[1]

P.1 $L(0) = R^n_+$, $0 \notin L(u)$ for $u > 0$.

P.2 $x \in L(u)$ and $x' \geqslant x$ imply $x' \in L(u)$.

P.3 If (a) $x > 0$, or (b) $x > 0$ and $(\bar{\lambda} \cdot x) \in L(\bar{u})$ for some $\bar{\lambda} > 0$ and $\bar{u} > 0$, the ray $\{\lambda \cdot x \mid \lambda \geqslant 0\}$ intersects $L(u)$ for all $u \in [0,\infty)$.

P.4 $u_2 \geqslant u_1 \geqslant 0$ implies $L(u_2) \subset L(u_1)$.

P.5 $\bigcap_{0 < u < u_0} L(u) = L(u_0)$ for $u_0 > 0$.

P.6 $\bigcap_{u \in [0,\infty)} L(u)$ is empty.

P.7 $L(u)$ is closed for all $u \in [0,\infty)$.

P.8 $L(u)$ is convex for all $u \in [0,\infty)$.

P.9 $E(u) = \{x \in L(u) \mid$ if $y \underset{\neq}{\leq} x$ *then* $y \notin L(u)\}$ is bounded for all $u \in [0,\infty)\}$.

From the description of the production technology and the axioms taken on the level sets, the *production, distance,* and *factor minimal cost functions* can be derived.

The Production Function

The production function $\Phi: R^n_+ \to R_+$ is defined by

$$\Phi(x) = \max\{u \mid x \in L(u)\}. \tag{3.18}$$

The production function Φ measures the largest output rate obtainable with x.[2]

[1] Shephard [1970a], p. 14.

[2] See Chapter 2, Shephard [1970] for the proof that Φ is well-defined.

The Distance Function

The distance function $\psi: R_+ \times R_+^n \to R_+$ is defined by

$$\psi(u,x) = \begin{cases} 0 & u > 0 \text{ and } \{\lambda \mid \lambda \cdot x \in L(u)\} = \varnothing \\ [\min\{\lambda \mid \lambda \cdot x \in L(u)\}]^{-1} & u > 0 \text{ and } \{\lambda \mid \lambda \cdot x \in L(u)\} \neq \varnothing \\ \infty & u = 0 \end{cases} \tag{3.19}$$

Essentially $\psi(u,x)$ measures how much one has to scale x, up or down, in order to just obtain output level u; that is, to put $\frac{x}{\psi(u,x)}$ on the boundary of $L(u)$ *if possible*. Observe that by the definition of $\psi(u,x)$ one has that

$$L(u) = \{x \mid \psi(u,x) \geqslant 1\}. \tag{3.20}$$

Hence, ψ is useful in that it may be used to characterize the level sets.[3]

The Factor Minimal Cost Function

Interpret $p \in R_+^n$ as prices and $p \cdot x$ as the cost of input vector x. If one defines the factor minimal cost function $Q: R_+ \times R_+^n \to R_+$ by

$$Q(u,p) = \min\{p \cdot x \mid x \in L(u)\}, \tag{3.21}$$

then $Q(u,p)$ would be the minimal cost of obtaining output level u.[4] The statement of Shephard's Duality begins to unfold when it is realized that $Q(u,p)$ has all of the properties that $\psi(u,x)$ has.[5] In effect, $Q(u,p)$ acts like a distance function. If it would be thought of as such, then the candidate "level sets" for which it would be a distance function would have to be

$$L_Q(u) = \{p \mid Q(u,p) \geqslant 1\}. \tag{3.22}$$

Thinking of $L_Q(u)$ as *price sets of the cost structure* much in the same way one thinks of $L(u)$ as

[3] Properties of ψ may be found in Chapter 3, Shephard [1970].

[4] As an easy consequence of the axioms $P.1$–$P.9$, $L(u) = E(u) + R_+^n = \overline{E(u)} + R_+^n$. Since $\overline{E(u)}$ is compact, it follows that $Q(u,p)$ is well-defined, i.e., the minimum is obtained.

[5] See Proposition 16, Shephard [1970].

the *production possibility sets of the production structure*, Shephard's Duality Theorem, loosely worded, is that we may derive one structure (cost or production) from the other. To quote Shephard (p.8, [1970]):

> "... production possibility sets of the production structure and the price sets of the cost structure are shown to be duals, derivable from each other by dual cost minimizations which determine the factor minimal and price minimal cost functions as dual distance functions."

To mathematically formulate the above statement, define the price minimal cost function as[6]

$$\psi^*(u,x) = \begin{cases} 0 & \text{if } L(u) = \varnothing \\ \inf\{p \cdot x \mid p \in L_Q(u)\} & \text{if } u > 0,\ L(u) \neq \varnothing \\ \infty & \text{if } u = 0 \end{cases} \tag{3.23}$$

Shephard's Duality Theorem is then the statement that

$$\Phi(x) = \max\{u \mid \psi^*(v,x) \geqslant 1\}. \tag{3.24}$$

3.2.2. Shephard's Approach to the Proof of Duality

Let $L^*(u) = \{x \mid \psi^*(u,x) \geqslant 1\}$. Shephard first shows that ψ^* had the properties of a distance function.[7] Next, he shows that the sets $L^*(u)$ defined a production technology, i.e., $L^*(u)$ satisfied axioms $P.1-P.9$.[8] Then, he shows that $L^*(u) = L(u)$ from which the duality statement may be easily derived and from which ψ^* may be seen to be equal to ψ.

The next two sections provide two alternate ways to proving Shephard's theorem. Version 1 we believe is far more direct and simpler to understand than the original approach. Version 2 motivates the definitions of ψ and ψ^* and in some sense *derives* the identity. It shows that to a

[6] Shephard defined ψ^* differently. There was no mention of the special case when $L(u) \neq \varnothing$ and $u > 0$.
[7] Proposition 39, Shephard [1970].
[8] Proposition 40, Shephard [1970].

large degree Shephard's theorem is a statement about supporting hyperplanes of boundary points of a particular class of convex subsets of R^n_+.

3.2.3. Proof of Shephard's Theorem: Version 1

The proof of Shephard's theorem is *immediate* if we can prove *first* that $\psi = \psi^*$. Note that if $\psi = \psi^*$ then ψ^* is a distance function, $L^* = L$, and L^* is a production technology.

Before we prove that $\psi = \psi^*$ we must first show that ψ^* is *well-defined* which in this case means that we must show that if $u > 0$, $L(u) \neq \emptyset$ then $L_Q(u)$ is nonempty. Since $L(u)$ is closed and convex with $0 \notin L(u)$ it follows by the well-known strong separation theorem for convex sets in Euclidean space that we can find a $p \in R^n$, $\alpha \in R$ such that $0 < \alpha < p \cdot z$ for all $z \in L(u)$.[9] Since $L(u)$ is monotonic it follows that $p \in R^n_+$. Dividing p by α produces an element in $L_Q(u)$. We are now ready to prove the theorem.

Shephard's Duality Theorem

The Distance Function equals the Price Minimal Cost Function, i.e., $\psi = \psi^*$.

Proof of Shephard's Duality Theorem (Version 1)

If either $u = 0$ or $u > 0$ but $L(u) = \emptyset$, then the result follows by definition. If we let $R(x) = \{\lambda \cdot x \mid \lambda \geqslant 0\}$ be the ray generated by $x \in R^n_+$, then only two possibilities are left to be considered:

(1) $u > 0$, $L(u) \neq \emptyset$, and $R(x) \cap L(u) = \emptyset$,

(2) $u > 0$, $L(u) \neq \emptyset$, and $R(x) \cap L(u) \neq \emptyset$.

Let us analyse case (1) first. By definition $\psi(u,x) = 0$. Hence we must show that $\psi^*(u,x) = 0$; that is, we must find a $p \in L_Q(u)$ such that $p \cdot x = 0$.

[9] See, for example, Bazarra and Shetty [1973], p.51.

Let $A(x) = \{i \mid x_i = 0\}$. Since $L(u)$ is monotonic, it is easy to see that $A(x)$ is non-empty. Set up the following optimization problem:

$$P: \min_{y \in L(u)} \sum_{i \in A(x)} y_i$$

Since $L(u) = \overline{E(u)} + R_+^n$ it follows that problem P is equivalent to problem P^*:

$$P^*: \min_{y \in \overline{E(u)}} \sum_{i \in A(x)} y_i .$$

$\overline{E(u)}$ was assumed compact; the objective function is continuous so we may conclude that a minimum is obtained. Let y^* denote such a minimum and δ denote the objective function value evaluated at y^*.

If $\delta = 0$, then for λ large enough $\lambda \cdot x \geqslant y^*$ which would imply that $\lambda \cdot x \in L(u)$. This in turn would mean that $R(x)$ meets $L(u)$ which is not the case we are considering. Hence, $\delta > 0$.

Define $p \in R_+^n$ by

$$p_i = \begin{cases} \frac{1}{\delta} & i \in A(x) \\ 0 & i \notin A(x) \end{cases} \tag{3.26}$$

Observe that $p \cdot x = 0$ and $p \cdot z \geqslant 1$ if $z \in L(u)$. Hence $p \in L_Q(u)$ with the desired properties. We have shown that $\psi^*(u,x) = 0$ as desired.

Let us now turn to case (2). In this case we have that $0 < \psi(u,x) < \infty$. For any $p \in L_Q(u)$ it is immediate that

$$p \cdot \frac{x}{\psi(u,x)} \geqslant 1 \tag{3.27}$$

since $\frac{x}{\psi(u,x)} \in L(u)$. Since (3.27) holds for any $p \in L_Q(u)$, by definition of $\psi^*(u,x)$ (and multiplying through by $\psi(u,x)$ in (3.27)) we have that $\psi^*(u,x) \geqslant \psi(u,x)$.

If $\psi^*(u,x) > \psi(u,x)$, then by definition of $\psi(u,x)$ (as a distance function) it must be the case that $\frac{x}{\psi^*(u,x)} \notin L(u)$. By the separation theorem we may find a $p \in R^n$ and an $\alpha \in R$ such that

$$p\cdot\frac{x}{\psi^*(u,x)} < \alpha < p\cdot z, \quad \forall z \in L(u). \tag{3.28}$$

As before we note that $p \in R^n_+$; thus $\alpha > 0$. By dividing p by α we have that $\frac{p}{\alpha}$ is in $L_Q(u)$ with $\frac{p}{\alpha}\cdot x < \psi^*(u,x)$ by (3.28). By taking the infimum over all $p \in L_Q(u)$ we would obtain the obvious contradiction that $\psi^*(u,x) < \psi^*(u,x)$. Hence $\psi^*(u,x) \leqslant \psi(u,x)$ which shows that $\psi^*(u,x) = \psi(u,x)$ in this case.

ur proof is complete.■

3.2.4. Proof of Shephard's Theorem: Version 2

Let S denote the set of all closed, convex, monotonic subsets of R^N_+ not containing the origin. Let $L \in S$. Let $R(x)$ denote the ray generated by an $x \in R^n_+$, i.e., $R(x) = \{\lambda\cdot x \mid \lambda \geqslant 0\}$. If $x \in R^n_+$, then either $R(x) \bigcap L = \varnothing$ or $R(x) \bigcap L \neq \varnothing$. If $R(x) \bigcap L \neq \varnothing$, then there exists a well-defined constant $\psi(x)$ such that $\frac{x}{\psi(x)}$ is on the boundary of L (relative to R^n_+). If $R(x) \bigcap L = \varnothing$, then define $\psi(x) = 0$.

Now pick an $x \in R^n_+$ and suppose $R(x) \bigcap L \neq \varnothing$. Let $\epsilon > 0$ and let $\overline{\epsilon}$ denote the vector $(\epsilon,\epsilon,\ldots,\epsilon) \in R^n_+$. Since ψ is continuous (Proposition (3.4.8), Appendix) then for some $\gamma > 0$ (to be determined later) we can find an $\epsilon > 0$ so that

$$\frac{x+\overline{\epsilon}}{\psi(x+\overline{\epsilon})} \in B_\gamma\left(\frac{x}{\psi(x)}\right) \tag{3.29}$$

($B_\gamma(\frac{x}{\psi(x)})$ denotes the open ball of radius γ about $\frac{x}{\psi(x)}$). Since $\frac{x+\overline{\epsilon}}{\psi(x+\overline{\epsilon})}$ is strictly positive it follows by an application of the separation theorem for convex subsets of R^n_+ that a

supporting hyperplane for L at $\dfrac{x+\overline{\epsilon}}{\psi(x+\overline{\epsilon})}$ exists. That is, there exists a $p \in R_+^n$, $p \neq 0$ such that

$$0 < p \cdot \frac{x+\overline{\epsilon}}{\psi(x+\overline{\epsilon})} \leqslant p \cdot z\,, \quad \forall z \in L\,. \tag{3.30}$$

Re-normalizing we may assume that $p \cdot \dfrac{x+\overline{\epsilon}}{\psi(x+\overline{\epsilon})} = 1$. It follows therefore that

$$1 = \sum_i p_i \cdot \frac{x_i+\epsilon}{\psi(x+\overline{\epsilon})} \geqslant \sum_{\{i\,|\,x_i>0\}} p_i \cdot \frac{x_i+\epsilon}{\psi(x+\overline{\epsilon})} \geqslant \left\{\sum_{\{i\,|\,x_i>0\}} p_i\right\}\left\{\frac{\min\limits_{\{i\,|\,x_i>0\}} x_i + \epsilon}{\psi(x+\overline{\epsilon})}\right\}.$$

Let

$$f(x,\epsilon) = \frac{\psi(x+\overline{\epsilon})}{\min\limits_{\{i\,|\,x_i>0\}} x_i + \epsilon}\,.$$

Then we have that

$$\sum_{\{i\,|\,x_i>0\}} p_i \leqslant f(x,\epsilon)\,. \tag{3.31}$$

In view of (3.29) and (3.31) we have that

$$p \cdot \left\{\frac{x}{\psi(x)} - \frac{x+\overline{\epsilon}}{\psi(x+\overline{\epsilon})}\right\} \leqslant \sum_{\{i\,|\,x_i>0\}} p_i \left\{\frac{x_i}{\psi(x)} - \frac{x_i+\epsilon}{\psi(x+\overline{\epsilon})}\right\} \leqslant f(x,\epsilon)\cdot\gamma\,. \tag{3.32}$$

Since $p \cdot \dfrac{x+\overline{\epsilon}}{\psi(x+\overline{\epsilon})} = 1$ rearranging (3.32) we obtain that

$$p \cdot x \leqslant \psi(x) + \psi(x) f(x,\epsilon)\gamma\,. \tag{3.33}$$

Now let

$$H^+ = \{(p,\alpha) \in R_+^n \times R_{++} \mid \exists x \in L \text{ such that } \alpha = p \cdot x \leqslant p \cdot z\,, \forall z \in L\}\,.$$

Since γ is arbitrary and $f(x,\epsilon)$ is bounded for ϵ small enough, it is easy to see that we have

just shown that

$$\inf_{\{(p,\alpha) \in H^+\}} \frac{p}{\alpha} \cdot x \leqslant \psi(x). \tag{3.34}$$

For $(p,\alpha) \in H^+$, $\frac{p}{\alpha} \cdot \frac{x}{\psi(x)} \geqslant 1$ so that, as an easy consequence,

$$\inf_{\{(p,\alpha) \in H^+\}} \frac{p}{\alpha} \cdot x \geqslant \psi(x). \tag{3.35}$$

Putting (3.34) together with (3.35) gives us

$$\inf_{\{(p,\alpha) \in H^+\}} \frac{p}{\alpha} \cdot x = \psi(x). \tag{3.36}$$

The interpretation of H^+ is precisely that it is the collection of supporting hyperplanes for L which separate the origin from L. By what we have just shown those points in the boundary which have supporting hyperplanes *in* H^+ are *dense* in the boundary. Further we have also shown that given any point z on the boundary there is a point y on the boundary arbitrarily close which has a supporting hyperplane $(p,\alpha) \in H^+$ whose "value" at z is arbitrarily close, i.e., $1 = \frac{p}{\alpha} \cdot y \geqslant \frac{p}{\alpha} \cdot z + \beta$, β small.[10]

We comment that it is *possible* that there may exist a point on the boundary which has *no* supporting hyperplane in H^+ to support L at this point. To see this, we construct the following example. Let $x_n = (1-\frac{1}{n}, 1-\frac{1}{n}, \frac{1}{2}^n)$ for $n \geqslant 1$ and let L be the smallest closed, convex, monotonic subset of R_+^3 containing the x_n's. Since $x_n \to (1,1,0)$, it follows that $(1,1,0) \in L$. We claim that no $(p,\alpha) \in H^+$ could exist which could support L at $(1,1,0)$. If there were such a $(p,\alpha) \in H^+$, then it is easy to see that (1) at least one of p_1 or p_2 must be positive, and (2) $p_3 > 0$.[11] Now we have that

[10] To see this, let $\epsilon > 0$, and $\gamma > 0$. Think of z as $\frac{x}{\psi(x)}$, y as $\frac{x+\bar{\gamma}}{\psi(x+\bar{\gamma})}$, and β as ϵ. Review equations (3.29) and (3.32).

[11] If (1) were not true, then $\alpha = 0$; if (2) were not true then $p \cdot x_n < p \cdot (1,1,0)$ for any n.

$$p\cdot x_n = (1-\frac{1}{n})(p_1+p_2) + \tfrac{1}{2}^n p_3 = (p_1+p_2) + \{\tfrac{1}{2}^n p_3 - \frac{1}{n}(p_1+p_2)\}. \tag{3.37}$$

Since $\frac{n}{2^n} \to 0$ as $n \to \infty$ we may find an n large enough so that $\tfrac{1}{2}^n p_3 - \frac{1}{n}(p_1+p_2) < 0$ implying that $p\cdot x_n < p\cdot(1,1,0)$, a contradiction.[12] We remark that this example clearly illustrates directly why the infimum is used in (3.36) as opposed to the minimum.

It is easy to see that since $(p,\alpha) \in H^+$ if and only if $(\frac{p}{\alpha},1) \in H^+$ we may identify H^+ with the set[13]

$$H^* = \{(p,1) \in R^n_+ \times R_{++} \mid \exists x \in L \text{ such that } 1 \leq p\cdot x \leq p\cdot z\,, \forall z \in L\}.$$

This last set of course may be identified with the set

$$\{p \in R^n_+ \mid \exists x \in L \text{ such that } 1 \leq p\cdot x \leq p\cdot z\,, \forall z \in L\}.$$

Thinking of L as a level set $L(u)$, this last set is precisely $L_Q(u)$ and $\psi(x)$ is really $\psi(u,x)$. *So looking at (3.36), we see that we have shown Shephard's Duality in the case identified in the previous section as case (2).*

So, by identifying $L_Q(u)$ with H^+ (which has already been interpreted) Shephard's Duality theorem in this special case reduces to a statement about supporting hyperplanes of closed, convex, monotonic subsets of R^n_+. In other words, loosely worded, the level set is characterized by the boundary which is characterized by the distance function on one hand (the boundary of L is the set of x which attains the "distance" of 1) and by the price minimal cost function on the other hand (which approximates the boundary with supporting hyperplanes).

The reason why ψ and ψ^* are defined differently for special cases is simply that one desires to extend the equality of ψ and ψ^* to the entire domain of R^n_+. Let us see how we may

[12] This can be done because the terms p_3 and (p_1+p_2) are both positive.

[13] This follows because p can be arbitrarily re-normalized. In one case we associate for $(p,\alpha) \in H^+$ the point $(\frac{p}{\alpha},1) \in H^*$. In the other case, we associate for $(p,1) \in H^*$ the point $(p,p\cdot x) \in H^+$ where x is a point of support in L for p. Rigorously, one could define equivalence classes of the form $[(p,1)]$ where $(p,\alpha)\sim(q,\beta)$ if and only if $\exists c > 0$ such that $(p,\alpha) = c(q,\beta)$.

motivate their respective definitions from this perspective.

First, suppose $u = 0$, thus $L(u) = R^n_+$. The idea of $\psi(u,x)$ is that $\frac{x}{\psi(u,x)}$ should be a point on the boundary of $L(u)$ on the ray generated by x "closest" to the origin. Since $0 \in L(0)$ a natural choice for $\psi(0,x)$ is to take it to be equal to ∞. Hence, we take $\psi^*(0,x) = \infty$.

Now suppose $x \in R^n_+$, $L(u) \neq \varnothing$ but $R(x) \cap L(u) = \varnothing$ (case (1)). From properties on convex subsets of R^n_+ we showed in the previous section that there is a $p \in R^n_+$ such that $\min\{p \cdot z \mid z \in L(u)\} = 1$ and $p \cdot x = 0$; that is, $\psi^*(u,x) = 0$ in this case. Hence, we take $\psi(u,x)$ to be 0 in this case.

Finally, suppose $L(u) = \varnothing$. In short, none of the functions are well-defined for this case. Since $L(u) = \varnothing$ implies that $R(x) \cap L(u) = \varnothing$ to be consistent, we therefore take *both* ψ and ψ^* to be 0 in this case.

The upshot of this development and proof of Version 2 is that Shephard's Duality Theorem may be motivated by studying properties of supporting hyperplanes of a particular class of convex subsets.[14] One does utilize heavily the simple and useful characterization of the boundary by the distance function. One obtains the Duality Statement as stated by Shephard when one interprets the functions in an economic light.

3.2.5. Extensions to the Dynamic Case

To extend the proof, as given in Version 1, of Shephard's Duality Statement for the finite horizon Dynamic Case we make some modifications. Instead of defining $A(x) = \{i \mid x_i = 0\}$ one defines $A_t(x) = \{t \mid x_i(t) = 0\}$. Instead of solving the optimization problem as given in the proof of case (1) we re-write it as

$$P': \min_{y \in LN(u)} \sum_i \int_0^h \{y^i \cdot 1_{A_i}\} d\mu .$$

[14] We remark that we "proved" that $\psi = \psi^*$ in this version by extending the definitions of ψ and ψ^*.

After making the appropriate modifications for the statement of Shephard's Duality in the Dynamic (Function Space) case the proof of Version 1 still applies because

(a) $\mathrm{LN}(u)$ is (weak-star) closed (see Theorem (3.1.1)).

(b) $\overline{\mathrm{EN}(u)}$ is (weak-star) compact (see Proposition (3.1.6)).

(c) $\mathrm{LN}(u) \subset \overline{\mathrm{EN}(u)} + (L_+^\infty(\mu))^n$ (see Theorem (3.1.1)).

(d) Separation Theorem for Locally Convex Topological Vector Spaces applies.

(We only proved (a)-(c) under the assumption of a finite horizon.)

We make two final remarks. First, to extend the proof of Shephard's Duality Statement, as given in Version 1, for the *infinite* horizon case additional hypotheses would have to be assumed to insure that $\mathrm{LN}(u)$ is (weak-star) compact. The axiomatic system presented in Section 2.2 is not strong enough to guarantee this property. Second, it is not possible to extend the proof of Shephard's Duality Statement, as given in Version 2, because our counter-example (Section 3.1.3) shows that strictly positive points need not be supportable.

3.3. On Laws of Diminishing Returns

In this section we show how our general framework of a production process may be used to deduce two versions of the Law of Diminishing Returns as formulated by Shephard. By proving such laws from the axiomatic framework, we reduce the question of whether such laws are "true" to the question of whether the axiomatic framework is reasonable. In this sence we hope to get "at the root" of such laws.

The two general types of laws of return posed by Shephard and considered here are: (1) Laws of Return for bounded input rates of Essential factors, and (2) Laws of Return for bounded intervals. The first type is defined and proved in Section 3.3.1. The second type is defined and proved in Section 3.3.2.

3.3.1. Laws of Return for Bounded Input Rates of Essential Factors

Taken from Shephard and Fare [1980], p.98 we describe this law in their own words:

> "A law of return so expressed for the static model of production is one of a law of bounded output rate. It is suggested for input and output rates which are not constant, i.e., for the dynamic structure of production, that a law of bounded output rate may hold, i.e., if time histories for *essential* factors are subject to an upper bound on input rate, the related output rate histories will be bounded in some way under unlimited increase in the maximal time rate of the input rate histories of the other factors."

We prove this statement, *solely in words*, from the axiomatic framework for the *finite* horizon case.

Proof of the First Type of the Law of Diminishing Returns

Suppose the vector of system exogenous inputs is norm bounded. As an easy consequence it follows that the cumulative amounts of the applications of exogenous resources to any activity must be bounded in magnitude (in the finite horizon). By Axiom 2 on the flows of

goods and services it follows that the flow types associated with the applications of exogenous resources is norm bounded. By Axiom 8 on the activity Production Functions this in turn means that the outputs realized through the production process for any activity must be norm bounded. As an easy consequence of this fact the cumulative amounts of the outputs transferred to the other activities from a given activity must also be bounded in magnitude. This means that the possible cumulative amounts of final outputs obtainable through production is bounded in magnitude hence in norm by Axiom 2.

If we restrict attention to just bounding in norm an essential set of system exogenous inputs, then in view of Axiom 8 the comments above still apply. Thus regardless of the magnitude of the other types of "nonessential" inputs the output rate would be bounded. And this is the statement of the first version of the Law (in the finite horizon case).

3.3.2. Laws of Return for Bounded Intervals of Essential Factor Application

Taken from Shephard and Fare [1980], p.100 we describe this law in their own words:

"The time spans over which essential input rates may be or are applied positively need not be infinite, that is the support of an input may be bounded, and unbounded time substitutions for resources may not be permitted. Then the question arises how outputs may be limited by limitations on the intervals of time over which essential factors are applied. Propositions of this type are laws of return for bounded intervals of application of essential factors."

To prove this law, we will first prove a stronger statement.

Proposition (3.3.1)

If $x \in \mathrm{LN}(u)$ such that for some $T > 0$, $\{x^j > 0\} \subset [0,T]$ for all j, then there exists some T^* such that $\{u^k > 0\} \subset [0,T^*]$ for all k.[1]

[1] It is sufficient to prove the Law for the Essential Set of inputs consisting of *all* inputs.

To pave the way for a "verbal" proof of Proposition (3.3.1), we prove a simple proposition below. For this proposition, let L denote a space associated with a particular flow type. Let ϵ denote the "lower bound" on the flows in L as dictated by Axiom 3. Let δ denote the "minimal spacing" between the set-up times of a flow in L as dictated by Axiom 4.

Lemma (3.3.2)

If $x,z \in L$ such that $\int_0^\tau x d\mu \leqslant \int_0^\tau z d\mu$ for all $\tau \in R_+$ and $\{z > 0\} \subset [0,T]$ for some T, then there exists a T^* such that $\{x > 0\} \subset [0,T^*]$.

Proof of Lemma (3.3.2)

If the cardinality of the number of set-up times in $S(x)$ (see 2.19) is finite, then it is easy to see that $T^* = sup\{t \mid t \in S(x)\}$ has the desired properties.

Suppose then that the cardinality of $S(x)$ is infinite. By Axiom 4 we may index $S(x)$ by a set of time points t_i, $i = 1,2, \ldots,$ such that

$$0 \equiv t_0 < t_1 < t_2 < \cdots \quad \text{such that } |t_i - t_{i-1}| \geqslant \delta, \quad \forall i. \tag{3.38}$$

By the definition of the t_i's it follows that either

$$\int_{[t_i,t_{i+1})} x d\mu = 0 \quad \text{or} \quad \int_{[t_i,t_{i+1})} x d\mu \geqslant \min\{\epsilon(t_{i+1} - t_i), \epsilon\},$$

the latter inequality due to Axiom 3.[2] But then it follows that

$$\int_{[t_i,t_{i+1})} x d\mu \geqslant \min\{\epsilon \cdot \delta, \epsilon\} \;\text{ if }\; \int_{[t_i,t_{i+1})} x d\mu \neq 0. \tag{3.39}$$

[2] We must account for the possibility that L is an event-based flow type. If this were not the case, then the t_i's would have to be time-grid points; so $\int_{[t_i,t_{i+1})} x d\mu = x(t_i)$ which is either 0 or at least ϵ.

Now let $N = \{i \mid \int_{[t_i,t_{i+1})} x d\mu \neq 0\}$. Since the cardinality of $S(x)$ is infinite it follows that the cardinality of N is infinite. But this could not be the case because we would have that

$$+\infty > \int_0^T z d\mu \geqslant \int_{R_+} x d\mu \geqslant \sum_{i=0}^{\infty} \int_{[t_i,t_{i+1})} x d\mu = \sum_{i \in N} \int_{[t_i,t_{i+1})} x d\mu \geqslant \sum_{i \in N} \min\{\epsilon \cdot \delta, \epsilon\} = +\infty$$

So the proof of Lemma (3.3.2) is complete.■

Proof of Proposition (3.3.1)

Suppose the conditions of the theorem held. Find a feasible flow for x to support output level u. By Lemma (3.3.2), we may easily deduce that the supports of the flow type associated with the applications of exogenous resources to the activities are finite. By Axiom 11, it follows that the realized output of production for each activity is restricted to this finite horizon. By Axiom 12, only a finite amount of output may be realized through production in this finite horizon; hence by Lemma (3.3.2) the supports of the flow types associated with the intermediate product flows are finite. By a final application of Lemma (3.3.2) the supports of the flows types associated with the final output variables *must be finite. That is, eventually output must stop* and this is what we wanted to show.■

Proof of the Second Type of the Law of Diminishing Returns

Since the supports of the flow types associated with the final output variables is finite, it follows immediately by Axiom 2 that such flows must be norm bounded. This is Shephard's version of the Second Type of the Law of Diminishing Returns.

3.4. Appendix

In this appendix, we prove the technical propositions used in the main body of the text. As notation, let L_E^k denote the subset of $L_+^\infty(\mu)$ which satisfies Axiom 2,3, and 4 for an event-based flow type, let L_C^k denote the subset of $L_+^\infty(\mu)$ which satisfies Axioms 2,3, and 4 for a continuous flow type, and let $L_E = L_E^2 \bigcap L_E^3 \bigcap L_E^4$, $L_C = L_C^2 \bigcap L_C^3 \bigcap L_C^4$. It is understood that these symbols are generic and stand for a particular flow type. Finally, throughout this section, $L_+^\infty(\mu)$ is topologized with the relative weak-star topology.

Proposition (3.4.1)

If $\{x_\alpha\}$ and $\{y_\alpha\}$ are nets in $L_+^\infty(\mu)$ such that $x_\alpha \leqslant y_\alpha$ for all α and if $x_\alpha \to x$, $y_\alpha \to y$, then $x \leqslant y$.

Proof of Proposition (3.4.1)

Suppose this is not true. This implies that we find a set A with positive, finite measure and an $\epsilon > 0$ such that the function x exceeds the function y by ϵ on the set A. Let $p = 1_A$. Clearly, $p \in L^1$. Let ψ_p denote the continuous linear functional induced from p. By the continuity of p, $\psi_p(x_\alpha) \to \psi_p(x)$, $\psi_p(y_\alpha) \to \psi_p(y)$, and $\psi_p(x_\alpha) \leqslant \psi_p(y_\alpha)$ for all α. Hence, $\psi_p(x) \leqslant \psi_p(y)$. But by definition of p, $\psi_p(x) - \psi_p(y) > \epsilon\mu(A) > 0$. The contradiction is readily apparent. ∎

Proposition (3.4.2)

L_E and L_C are each closed in $L_+^\infty(\mu)$.

Before we prove Proposition (3.4.2), we first show that neither L_C^3 nor L_C^4 is by itself closed as the following two counter-examples show.

Counter-Example to the Closure of L_C^3

Let ϵ denote the lower bound described in Axiom 3. Let the time grid T (as described in Axiom 1) be the set of natural numbers. Define for $n \geqslant 1$,

$$x_{2n} = \epsilon \sum_{i=1}^{n} 1_{\left(\frac{i}{n}-\frac{1}{2n},\frac{i}{n}\right)}.$$

It may be verified that

$$x_{2n} \to \frac{\epsilon}{2} 1_{(0,1)}.$$

Of course, the function $\frac{\epsilon}{2} 1_{(0,1)}$ violates Axiom 3.

Counter-Example to the Closure of L_C^4

Let δ denote the minimal distance between set-up times as described in Axiom 4. Fix $h > 0$ and choose $K > \frac{h}{\delta}$. Again, let T be the set of natural numbers. Define for $n \geqslant 1$,

$$x_n = \sum_{i=1}^{K} 1_{\left(\frac{i}{K}-\frac{1}{2K},\frac{i}{K}\right)} + \sum_{i=1}^{K} \frac{1}{n} 1_{\left(\frac{i-1}{K}-\frac{1}{2K},\frac{i-1}{K}\right)}.$$

The set of set-up times for each x_n, $S(x_n)$, is clearly empty; thus, each x_n satisfies Axiom 4. But, clearly,

$$x_n \to \sum_{i=1}^{K} 1_{\left(\frac{i}{K}-\frac{1}{2K},\frac{i}{K}\right)},$$

which has K set-up times in the horizon $[0,h]$. Thus, $\sum_{i=1}^{K} 1_{\left(\frac{i}{K}-\frac{1}{2K},\frac{i}{K}\right)}$ violates Axiom 4.

Proof of Proposition (3.4.2)

In light of the two counter-examples, we will prove Proposition (3.4.2) by first proving that L_C^3 (and L_E^3) is closed (Proposition 3.4.3) and then prove that $L_C^3 \cap L_C^4$ (and $L_E^3 \cap L_E^4$) is closed (Proposition 3.4.4).

Proposition (3.4.3)

L_C^2 and L_E^2 are each closed in $L_+^\infty(\mu)$.

Proof of Proposition (3.4.3)

By the continuity of parameters $\{h_k^i\}$ used to define Axiom 2 for an event-based flow type, it is immediate that L_E^2 is closed in $L_+^\infty(\mu)$. To show the closure of L_C^2, we will show that the complement of L_C^2 in $L_+^\infty(\mu)$ is open.

Pick a z in the complement of L_C^2 in $L_+^\infty(\mu)$. It follows by definition of Axiom 2 that for some $A \in R_+$ and I_k that

$$\int z \cdot 1_{I_k} d\mu \leqslant A \quad \text{but } \|z \cdot 1_{I_k}\|_\infty > g_k(A). \tag{3.40}$$

Let $\delta = \|z \cdot 1_{I_k}\|_\infty - g_k(A)$. Since g_k is continuous, we may find an $\epsilon > 0$ such that

$$|g_k(B) - g_k(A)| < \frac{\delta}{4} \quad \text{whenever } |B - A| < \epsilon. \tag{3.41}$$

Now define

$$p = 1_{[(z > \|z \cdot 1_{I_k}\|_\infty - \frac{\delta}{4}) \cap I_k]}$$

$$\alpha = (\int p d\mu) \frac{\delta}{4}$$

$$N = (z + N(p,\alpha)) \cap (z + N(1_{I_k},\epsilon)) \cap L_+^\infty(\mu)$$

where for $q \in L^1(\mu)$, $\beta > 0$

$$z + N(q,\beta) = \left\{ y \in L_+^\infty(\mu) \,\middle|\, \left|\int q(z-y) d\mu\right| < \beta \right\}.$$

(N is a neighborhood about z in this topology.)

Pick any $y \in N$. Let $B = \int y \cdot 1_{I_k} d\mu$. Since $y \in z + N(p,\alpha)$ it is easy to verify that

$$\|y \cdot 1_{I_k}\|_\infty > \|z \cdot 1_{I_k}\|_\infty - \frac{\delta}{2}.$$

By (3.40) and (3.41), we then have that

$$\|y \cdot 1_{I_k}\|_\infty > \|z \cdot 1_{I_k}\|_\infty - \frac{\delta}{4} = g_k(A) + \frac{\delta}{2}. \tag{3.42}$$

There are two cases two consider.

Case 1: $|B - A| \leqslant \epsilon$.

Since $|B - A| \leqslant \epsilon$, by (3.41) and (3.42), it is immediate that

$$\|y \cdot 1_{I_k}\|_\infty \geqslant g_k(B) + \frac{\delta}{4}$$

which implies that y is in the complement of L_C^2 in $L_+^\infty(\mu)$.

Case 2: $|B - A| > \epsilon$.

Since y is also in $z + N(1_{I_k}, \epsilon)$, we have that

$$\int z \cdot 1_{I_k} d\mu - \epsilon \leqslant \int y \cdot 1_{I_k} d\mu \leqslant \int z \cdot 1_{I_k} d\mu + \epsilon \leqslant A + \epsilon.$$

Hence, $B \leqslant A + \epsilon$. Since $|B - A| > \epsilon$, it follows that $B < A - \epsilon$. By monotonicity of g_k, we then have that

$$g_k(B) \leqslant g_k(A - \epsilon) \leqslant g_k(A). \tag{3.43}$$

It is immediate by (3.42) and (3.43) that y is in the complement of L_C^2 in $L_+^\infty(\mu)$.

In either case 1 or case 2 we have shown that if $y \in N$, then y is in the complement of L_C^2 in $L_+^\infty(\mu)$. Thus, the complement of L_C^2 in $L_+^\infty(\mu)$ is open proving the desired result. ■

Proposition (3.4.4)

$L_E^3 \bigcap L_E^4$ and $L_C^3 \bigcap L_C^4$ are each closed in $L_+^\infty(\mu)$.

Proof of Proposition (3.4.4)

By the definition of an event-based, flow type and the properties imposed on the time grid T, it is easily verified that L_E^3 and L_E^4 are each closed in $L_+^\infty(\mu)$. Hence, so is $L_E^3 \bigcap L_E^4$.

To facilitate the proof of closure of $L_C^3 \bigcap L_C^4$, we first prove a simple lemma.

Lemma (3.4.5)

Let $x \in L_C$ and let $S(x)$ denote the set of set-up times for x (see 2.19). If $I = [a,b]$ such that $\lambda\{I \bigcap (x=0)\} > 0$ and $\lambda\{I \bigcap (x>0)\} > 0$, then $S(x) \bigcap I \neq \emptyset$.

Proof of Lemma (3.4.5)

If $a \in S(x)$, then the result follows immediately. So, suppose $a \notin S(x)$. By definition of $S(x)$, there must be an interval of the form $[a,c]$, $c > a$, such that either $\lambda\{[a,c] \bigcap (x=0)\} = 0$ or $\lambda\{[a,c] \bigcap (x>0)\} = 0$. Without loss of generality assume that $\lambda\{[a,c] \bigcap (x>0)\} = 0$. Let $t^* = sup\{d \in I \mid \lambda\{[a,d] \bigcap (x>0)\} = 0\}$.

We now claim that $t^* \in S(x)$. If $t^* \notin S(x)$, then by definition of $S(x)$ we could either find a $t^{**} > t^*$, $t^{**} \in [a,b]$, which has the property that $\lambda\{[a,t^{**}] \bigcap (x>0)\} = 0$ or t^* must equal b. By definition of t^*, it must be the second case, i.e., $t^* = b$. But then this means that $\lambda\{I \bigcap (x>0)\} = 0$ which violates our original assumption. The result follows.■

We resume the proof of Proposition (3.4.4). Let $\{x_\alpha\}$ be a net in $L_C^3 \bigcap L_C^4$ converging to some x. Suppose $x \notin L_C^3$. Then $\{0 < x < \epsilon\}$ has positive Lebesgue measure. Since Lebesgue measure is regular, it follows that there is a finite interval $[a,b] \subset \{0 < x < \epsilon\}$. Choose a $t \in (a,b)$. Let δ denote the minimal distance between set-up times as described in Axiom 4. Choose $\epsilon \in (0,\frac{\delta}{3})$ so that if $I_1 = [t-\frac{\epsilon}{2}, t+\frac{\epsilon}{2}]$ and $I_2 = [t+\epsilon, t+2\epsilon]$ then $I_j \subset (a,b)$, $j = 1,2$.

We make the following claim: for some β, $S(x_\alpha) \bigcap I_j \neq \varnothing$, $j=1,2$, for all $\alpha \geqslant \beta$. If this were not true, then we could extract a countable sub-net $\{x_n\}$ such that $S(x_n) \bigcap I_j = \varnothing$, $j=1,2$. By Lemma (3.4.5), it follows that for each n either $j=1,2$. This in turn would imply that either $\int x_n \cdot 1_{I_j} d\mu = 0$ or $\int x_n \cdot 1_{I_j} d\mu \geqslant \epsilon\lambda(I_j) > \int x \cdot 1_{I_j} d\mu > 0$, $j=1,2$. For $p^j = 1_{I_j}$, $j=1,2$, we would then have that x_n does not converge to x--an obvious contradiction.

So we have shown that for some β, $S(x_\alpha) \bigcap I_j \neq \varnothing$, $j=1,2$, for all $\alpha \geqslant \beta$. By construction of each I_j it is immediate then that, for $\alpha \geqslant \beta$, x_α violates Axiom 4. And this is a contradiction. Hence, x must be in L_C^3.

Suppose $x \notin L_C^4$. Then we may find a t_1 and t_2 each in $S(x)$ such that $|t_2 - t_1| < \delta$. Choose $\epsilon \in (0, \frac{\delta - |t_2 - t_1|}{2})$ so that if $I_j = [t_j - \epsilon, t_j + \epsilon]$, $j=1,2$, then $I_1 \bigcap I_2 = \varnothing$. By definition of $S(x)$, we have that $\lambda\{I_j \bigcap (x>0)\} > 0$, $j=1,2$. Since $x_\alpha \to x$, we then have that, for all α,

$$\int x_\alpha \cdot 1_{\{I_j \bigcap (x=0)\}} d\mu \to \int x \cdot 1_{\{I_j \bigcap (x=0)\}} = 0, \quad j=1,2 \tag{3.44}$$

$$\int x_\alpha \cdot 1_{\{I_j \bigcap (x>0)\}} d\mu \to \int x \cdot 1_{\{I_j \bigcap (x>0)\}} \geqslant \epsilon\lambda\{I_j \bigcap (x>0)\} > 0, \quad j=1,2. \tag{3.45}$$

(We are using the fact that $x \in L_C^3$ in (3.45).) By (3.44) and the fact that each $x_\alpha \in L_C^3$ it follows that for some β_j, $j=1,2$, $\lambda\{I_j \bigcap (x_\alpha = 0)\} > 0$ for all $\alpha \geqslant \beta_j$, $j=1,2$. Similarly, by (3.45) and the fact that each $x_\alpha \in L_C^3$, it follows that for some γ_j, $j=1,2$, $\lambda\{I_j \bigcap (x_\alpha > 0)\} > 0$ for all $\alpha \geqslant \gamma_j$, $j=1,2$. Thus, by Lemma (3.4.5) and the above, we have that, for all $\alpha \geqslant \max_{j=1,2}\{\beta_j, \gamma_j\}$, $S(x_\alpha) \bigcap I_j \neq \varnothing$, $j=1,2$. This implies that $x_\alpha \notin L_C^4$ which is an obvious contradiction. Hence, x must be in L_C^4.

The above argument shows that $x \in L_C^3 \bigcap L_C^4$. The result follows.■

Proposition (3.4.6)

Let $\{x_\alpha\}$, $\{y_\alpha\}$ be nets in L_C such that, for all α, $\lambda\{(x_\alpha > 0) \bigcap (y_\alpha >)\} = 0$. If $x_\alpha \to x$ and $y_\alpha \to y$, then $\lambda\{(x > 0) \bigcap (y > 0)\} = 0$.

In the discussion on Axiom 5 (closure of the domain), Section 2.2.2.1 we remarked that there may be additional constraints linking the domains of the inputs applied in production. One example we gave is when an activity utilizes one machine to produce two similar types of products. For this example, the functions which define the rate of machine hours applied to each type of product are linked in that both functions cannot be positive at the same time. It is desirable to show that this property preserves closure so as to give a plausible basis for accepting Axiom 5.

Before proving Proposition (3.4.6), we remark that if the nets $\{x_\alpha\}$ and $\{y_\alpha\}$ were only assumed to belong to $L_+^\infty(\mu)$ then Proposition (3.4.6) would be false. To see this, define for $n \geq 1$

$$x_{2n} = \sum_{i=1}^{n} 1_{\left(\frac{i}{n}-\frac{1}{2n}, \frac{i}{n}\right)}$$

$$y_{2n} = 1_{(0,1)} - x_{2n} .$$

It may be verified that

$$x_{2n} \to 1_{(0,1)} \text{ and } y_{2n} \to 1_{(0,1)}$$

and thus clearly the nets $\{x_{2n}\}$, $\{y_{2n}\}$ satisfy the conditions of Proposition (3.4.6) but violate the conclusion.

Proof of Proposition (3.4.6)

Suppose $\lambda\{(x > 0) \bigcap (y > 0)\} > 0$. Find a finite interval $[a,b] \subset \{(x > 0) \bigcap (y > 0)\}$. Choose a $t \in (a,b)$. Let δ denote the minimal distance between set-up times as described in Axiom 4. Choose $\epsilon \in (0,\frac{\delta}{3})$ so that if $I_1 = [t - \frac{\epsilon}{2}, t + \frac{\epsilon}{2}]$, $I_2 = [t+\epsilon, t+2\epsilon]$ then $I_j \subset [a,b]$,

$j=1,2$.

Since $x_\alpha \to x$, $x_\alpha \in L_C^3$ for each α, and $\lambda\{(x>0) \bigcap I_j\} > 0$, $j=1,2$, it follows easily that eventually, for some β_j, $\lambda\{(x_\alpha > 0) \bigcap I_j\} > 0$, $\forall \alpha \geqslant \beta_j$, $j=1,2$. Since $y_\alpha \to y$, $y_\alpha \in L_C^3$ for each α, and $\lambda\{(y_\alpha > 0) \bigcap I_j\} > 0$ $j=1,2$, it follows again that eventually, for some δ_j, $j=1,2$, $\lambda\{(y_\alpha > 0) \bigcap I_j\} > 0$, $\forall \alpha \geqslant \delta_j$, $j=1,2$. Since, for all α, $\lambda\{(x_\alpha > 0) \bigcap (y_\alpha > 0)\} = 0$ we have that for $\alpha \geqslant \delta_j$, $j=1,2$, $\lambda\{(x_\alpha > 0) \bigcap I_j\} > 0$. By Lemma (3.4.5), $S(x_\alpha) \bigcap I_j \neq \varnothing$, $j=1,2$. But this implies that for $\alpha \geqslant \max_{j=1,2}\{\beta_j, \delta_j\}$, $x_\alpha \notin L_C^4$ by our construction of I_j, $j=1,2$. This contradiction proves the result.■

Proposition (3.4.7)

Let $F: L_+^\infty(\mu) \to L_+^\infty(\mu)$ such that for each $x \in L_+^\infty(\mu)$, F is bounded (in norm) in a neighborhood of x. Assume further that for each $h \in R_+$, F satisfies the following property: $\forall$ $\epsilon > 0$,

$$\exists \delta > 0 \text{ such that } \left| \int_{[0,h]} (F(x)-F(y))d\mu \right| < \epsilon \text{ whenever } \left| \int_{[0,h]} (x-y)d\mu \right| < \delta .$$

Then F is weak-star continuous.

Proof of Proposition (3.4.7)

Pick an $x \in L_+^\infty(\mu)$. Let $U(x)$ denote the neighborhood about x in which F is bounded. Let $B(x)$ denote the bound. Let $F(x) + N(p^1, \ldots, p^n; \epsilon)$ denote a basic open neighborhood about $F(x)$. Let $\delta = \frac{\epsilon}{6} B(x)$. Find h large enough so that if $I = [h, \infty)$ then $\|p^i \cdot 1_I\|_1 < \delta$ for $i=1,2,\ldots,n$. Find integrable, simple functions q^i, $i=1,2,\ldots,n$ such that $\|p^i - q^i\|_1 < \delta$ for $i=1,2,\ldots,n$. Finally, let $\rho = \min_i \frac{\epsilon}{3\|q^i\|_1}$.

We may find a $\delta > 0$ that the property on F holds when $\epsilon = \rho$. Let $V(x) = y \in (x + N(1_{I_1}, \delta)) \bigcap (U(x)) \bigcap L_+^\infty(\mu)$. For any p_i, $i=1,2 \ldots, n$, we have that

$$\|p^i(F(x)-F(y))\|_1 \leqslant \|p^i(F(x)-F(y))\cdot 1_I\|_1 + 2B(x)\|p^i\cdot 1_I\|_1$$

$$\leqslant \|p^i-q^i\|_1\|F(x)-F(y)\|_\infty + \|q^i\|_\infty \cdot \|(F(x)-F(y))\cdot 1_I\|_1 + 2B(x)\|p^i\cdot 1_I\|_1$$

$$\leqslant (\frac{\epsilon}{6B(x)})B(x) + \|q^i\|_\infty(\frac{\epsilon}{3\|q^i\|_\infty}) + 2B(x)(\frac{\epsilon}{6B(x)}) = \epsilon .$$

Hence, we have found a neighborhood about x, $V(x)$, such that if $y \in V(x)$, $F(y) \in F(x) + N(p^1, \ldots, p^n; \epsilon)$. Thus, F is weak-star continuous.■

Proposition (3.4.8)

Let L denote a closed, monotonic *but not necessarily convex* proper subset of R^n_+ which is non-empty. Let $\psi(x)$ denote the distance function (dropping the u in $\psi(u,x)$). ψ is continuous on R^n_+.[1]

Before we prove this proposition, we remark that the proof of Shephard's Duality given in Section (3.2.4) utilized the fact that the distance function is continuous. Shephard [1970] proves that $\psi(u,x)$ is continuous on $R_+ \times R^n_+$ by first citing a theorem that states that a convex function defined on a convex open subset of R^n is continous on this open subset and then proving that ψ is both upper and lower semi-continous on the boundary.[2] Our proof below is simpler and more direct; furthermore, we do not assume that the level set is convex.

Proof of Proposition (3.4.8)

To prove this proposition, we first prove a very simple lemma.

Lemma (3.4.9)

[1] Since ψ is constant and hence continuous when $L = R^n_+$ (corresponding to the case $u = 0$) or when $L = \emptyset$ we omit these cases.

[2] Proposition 16, Shephard [1970]. Berge [1963], p. 193.

if $x \geqslant y$ then $\psi(x) \geqslant \psi(y)$.

Proof of Lemma (3.4.9)

If this were not true, then for some $x \geqslant y$ $\psi(x) < \psi(y)$. As $\psi(y) > 0$ this means that $R(y) \bigcap L \neq \varnothing$. Clearly this implies that $R(x) \bigcap L \neq \varnothing$ since $x \geqslant y$. By monotonicity of L and by definition of ψ we have that $\frac{x}{\psi(y)} \in L$ with $\frac{1}{\psi(y)} < \frac{1}{\psi(x)}$. This contradicts the definition of $\psi(x)$. The result follows.■

We now proceed with the proof of Proposition (3.4.8). First we show that ψ is continuous on $\{x \mid \psi(x) = 0\}$. Let x be such that $\psi(x) = 0$. Define y^n by $y_i^n = x_i + \frac{1}{n}$ for each i.

Claim (3.4.10)

$\psi(y^n) \to 0$ as $n \to \infty$.

Proof of Claim (3.4.10)

If this were not true, then for some $\epsilon > 0$ we could extract a subsequence y^{n_k} such that $\psi(y^{n_k}) \geqslant \epsilon$ for all k. By Lemma (3.4.9) it follows that, for each k, $\epsilon \leqslant \psi(y^{n_k}) \leqslant \psi(y^{n_1})$. Thus $\{\psi(y^{n_k})\}$ is contained in a compact subset and therefore has a convergent subsequence. (We will not change notation for the subsequence.)

Hence for some $\rho > 0$ we have that

$$\frac{y^{n_k}}{\psi(y^{n_k})} \to \frac{x}{\rho}. \tag{3.46}$$

But $\frac{y^{n_k}}{\psi(y^{n_k})} \in L$ for each k and since L is closed this means that $\frac{x}{\rho} \in L$ too. But then $R(x)$ would meet L implying that $\psi(x) > 0$, a contradiction. The result follows.■

Now let $y \in B_{\frac{1}{n}}(x) \bigcap R_+^n$. By Lemma (3.4.9) we have that $\psi(y) \leqslant \psi(y^n)$ for each n. By Claim (3.4.10), it is now easy to deduce that ψ is continuous on $\{x \mid \psi(x) \leqslant 0\}$.

Let us now turn to proving that ψ is continuous on $\{x \mid \psi(x) > 0\}$. Define z^n by

$$z_i^n = \begin{cases} x_i - \dfrac{1}{n} & \text{if } x_i > 0 \\ 0 & \textit{otherwise} \end{cases}$$

To faciltate the proof, we first make two simple claims.

Claim (3.4.11)

$\forall \alpha > 0$ eventually $\psi(y^n) < \psi(x)(1+\alpha)$.

Proof of Claim (3.4.11)

As before, if this were not true we could extract a convergent subsequence such that for some $\rho > 0$

$$\frac{y^{n_k}}{\psi(y^{n_k})} \to \frac{x}{\rho} \tag{i}$$

$$\frac{1}{\rho} \leqslant \frac{1}{\psi(x)(1+\alpha)} < \frac{1}{\psi(x)} \tag{ii}$$

Again, $\frac{x}{\rho} \in L$. Since $\frac{1}{\rho} < \frac{1}{\psi(x)}$ we contradict the definition of $\psi(x)$. The result follows.■

Claim (3.4.12)

$\forall \alpha > 0$ eventually $\psi(z^n) > \psi(x)((1-\alpha)$.

Proof of Claim (3.4.12)

If this were not true, then we could find a subsequence $\{z^{n_k}\}_{k=1}^{\infty}$ such that $\psi(z^{n_k}) \leqslant \psi(x)(1-\alpha)$ for each k. It is a simple general fact that if $y \leqslant x$ with $y_i = 0$ if and only

if $x_i = 0$ then $R(x) \bigcap L \neq \emptyset$ implies that $R(y) \bigcap L \neq \emptyset$. From this we can *now* say that

$$0 < \psi(z^{n_1}) \leqslant \psi(z^{n_k}) \leqslant \psi(x)(1-\alpha), \quad \forall k.$$

As before, extract a convergent subsequence so that for some $\rho > 0$ we have that

$$\frac{z^{n_k}}{\psi(z^{n_k})} \to \frac{x}{\rho} > \frac{x}{\psi(x)}. \tag{3.47}$$

Since $\frac{z^{n_k}}{\psi(z^{n_k})}$ is in the boundary of L for each k (which is closed) then $\frac{x}{\rho}$ must be in the boundary. But by (3.48) this is clearly not the case. A contradiction is reached thus proving the claim.■

Let $\alpha > 0$. By Claims (3.4.11) and (3.4.12), we can find $n(\alpha), m(\alpha)$ so that $\psi(y^{n(\alpha)}) < \psi(x)(1+\alpha)$ and $\psi(z^{m(\alpha)}) > \psi(x)(1-\alpha)$. Let $N(\alpha) = \max\{n(\alpha), m(\alpha)\}$. Let $h \in B_{\frac{1}{N(\alpha)}}(x) \bigcap R_+^n$. By Lemma (3.4.9) we have that

$$\psi(x)(1-\alpha) < \psi(z^{m(\alpha)}) \leqslant \psi(h) \leqslant \psi(y^{N(\alpha)}) < \psi(x)(1+\alpha)$$

Therefore $|\psi(h) - \psi(x)| < \alpha\psi(x)$. As α is arbitrary, it follows that ψ is continuous on $\{x \mid \psi(x) > 0\}$. Hence, ψ is continuous.■.

4. APPLICATION OF THE GENERAL MODEL TO MULTI-PROJECT RESOURCE-USE PLANNING

In this chapter, we use the general model as a tool to provide a systematic analysis of a heuristic solution proposed by Leachman and Boysen [1983] for the problem of *multi-project resource-use planning* for a *multi-project production system.* The problem is to determine explicit resource allocations through time to projects to insure that schedules are met. Our systematic analysis not only provides a logical foundation for their approach but also shows how their approach can be extended and improved. More importantly, the analyses carried out in this chapter illustrate the value of using a general conceptual framework of a production system to evaluate proposed heuristic solutions to production planning problems.

4.0. Introduction

A *multi-project production system* G is a production system comprised of a number of single-project production systems $G_1, G_2, \ldots, G_k$ each utilizing the same set of system exogenous inputs. In these organizations, project managers are responsible for keeping projects on schedule and within budget. Without an effective method for allocating scarce resources to the projects frequent project delays ensue. Such delays could be avoided if an effective method for *multi-project resource-use planning* were available; that is, a method for determining explicit resource allocations to projects through time to insure that schedules were met.

Traditionally, decision support systems for multi-project planning develop resource-constrained schedules of the activities within each single-project production system by treating the multi-project system as if it were one, large single-project system.[1] In our model of a single-project production system (see Section 2.3.4), once the schedules for the activities are determined the resource allocations for the activities and hence the resource allocations to the projects are determined. The problem with this approach is twofold: first, developing resource-constrained scheduling of the activities for *large* single-project production systems is

[1] See, for example, Kurtulus and Davis [1982].

computationally unattractive; second, higher-level management is not responsible for detailed scheduling of individual activities. For similar reasons, it is also unsatisfactory to approach the problem by determining resource allocations to the activities. A new approach is needed.

Leachman and Boysen in their 1983 paper, "An Aggregate Model for Multi-Project Resource Allocation," present a novel approach to solving the problem of multi-project resource-use planning.[2] The general idea is to combine activities in each single-project production system into "aggregate activities" and then allocate resources to the aggregate activities. It is envisioned that the allocations of resources to the aggregate activities would serve as resource constraints for the resource-constrained scheduling to follow. For the method to work the choice of allocations to aggregate activities must insure that the set of schedules consistent with the allocations includes schedules which are on time and within budget.

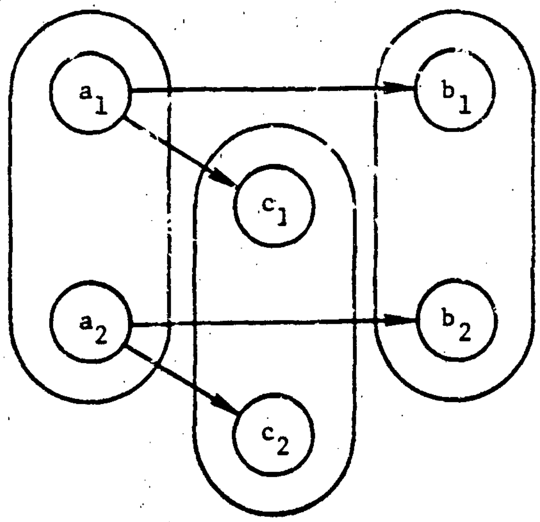
To develop the general idea of aggregation, let G denote a production network associated with a single-project production system. When activities are combined into aggregate activities, an "aggregate production network" G' derived through structural aggregation from G is created.

Definition (4.1)

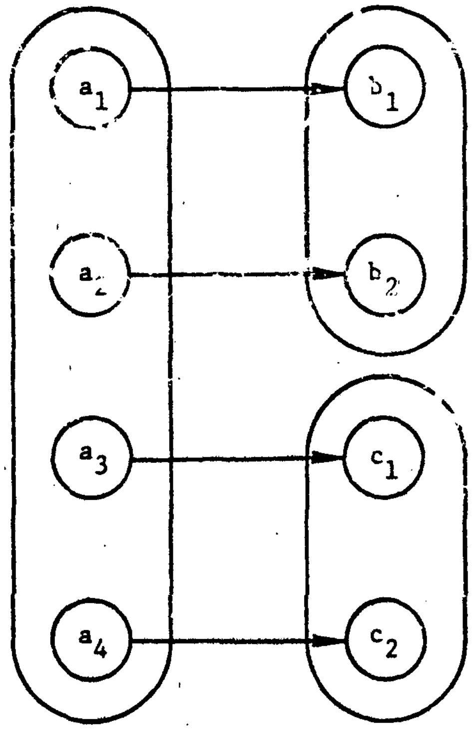
A network G' with M nodes is said to be *derived through structural aggregation* from a network G with N nodes if G' is acyclic, directed, and with the following property: there exists a surjective map ψ: $\{1,2,\ldots,N\} \rightarrow \{1,2,\ldots,M\}$ such that arc (i,j) is in G' if and only if there exists an arc (k,l) in G with $k \in \psi^{-1}(i)$, $l \in \psi^{-1}(j)$.

For example, Figures (4-1a) and (4-1b) show two detailed subnetworks G_1 and G_2 associated with a single-project production system. The circles indicate which activities have been combined into aggregate activities. The resulting aggregate subnetworks G'_1 and G'_2 of the aggregate production networks derived through a structural aggregation are shown in Figures (4-1c) and (4-1d). Note that Figures (4-1c) and (4-1d) are identical. This example shows that two

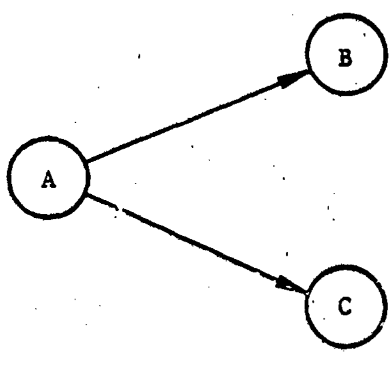
[2] See also Leachman and Boysen [1982].



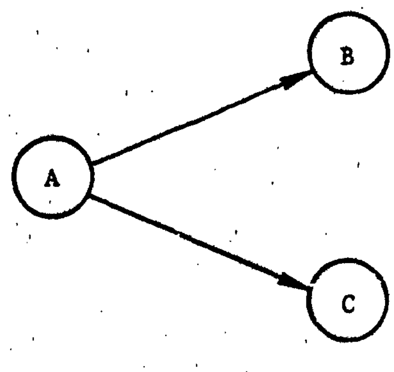
(a) Detailed Subnetwork G_1



(b) Detailed Subnetwork G_2



(c) Aggregate Subnetwork G_1'



(d) Aggregate Subnetwork G_2'

FIGURE (4-1)

EXAMPLES OF STRUCTURE UNDERLYING AGGREGATE SUBNETWORKS

aggregate production networks may be identical but the underlying production networks from which they were derived may be fundamentally different.

For this example, it is intuitively clear that the choice for resource allocations to aggregate activities B and C (in G'_1 or G'_2) should be dependent in some way on the choice for the resource allocations to aggregate A. Since G_1 is fundamentally different from G_2, it is also intuitively clear that the dependence between choices for resource allocations of aggregate activities A, B, and C in G'_1 is different from the dependence between choices for resource allocations of aggregate activities A, B, and C in G'_2. Essentially, Leachman and Boysen's method for determining resource allocations to aggregate activities to facilitate scheduling of the activities *is* a method for *modeling* the dependence between the choices for resource allocations to aggregate activities. The example hopefully motivates why the production network and the aggregate production network must *both* be considered when developing a method for modeling the dependence between choices for resource allocations to aggregate activities.

In their paper, Leachman and Boysen give an example of a subnetwork G_1 of G and a subnetwork G'_1 derived through structural aggregation from G_1 for which it would be difficult to model the "dependence relationships" between the aggregate activities in G'_1 so as to facilitate the detailed scheduling. Hence, they present only those subnetworks G_1 and G'_1 for which they could model the dependence relationships using linear constraints. They then formulated a linear program to accomplish the multi-project resource-use planning.

The ideas presented in Leachman and Boysen's paper are novel, innovative, and intuitively appealing. They recognized that dependence relationships between aggregate activities exist and that modeling these dependence relationships is intimately related to the structure of the subnetworks at both levels. The fundamental problem with their approach is that is does not present a methodology for attacking the general problem of how one should model the dependence relationships for an arbitrary subnetwork. Furthermore, certain modeling techniques were employed but were not adequately justified.

In this chapter we use the general model as a tool to provide a *systematic* approach to analyzing how one should model the dependence relationships. To a lesser extent, our systematic approach provides a mathematical basis for accepting some of their modeling techniques. The approach, being structured and systematic, enhances clarity and thus provides valuable insight into their approach. To a larger extent, our systematic approach provides the means for modeling dependence relationships for a larger class of subnetworks. Furthermore, for one important class of subnetworks, our approach differs substantially from theirs. For this class of subnetworks, we feel our approach is more sensible.

The key idea to developing a systematic approach to modeling the dependence relationships between aggregate activities is to realize that Leachman and Boysen's approach for multi-project resource-use planning is a production planning technique for an "aggregate" production system (albeit, a conceptual one). As mentioned in Chapter 1, to do production planning for an (aggregate) production system the input-to-output transformation must be modeled for the (aggregate) production system. Our systematic approach begins by assuming that the correspondence which defines the input-to-output transformation for the aggregate production system satisfies the axioms in Chapter 2. This assumption implies that to model the input-to-output transformation the flow types associated with the aggregate production system must be modeled (i.e., one must model the aggregate production functions, the intermediate product transfers between aggregate activities, and the applications of system exogenous and intermediate product inputs to the aggregate activities). Once the flow types have been modeled, the set Z of feasible choices for the allocations of resources to the aggregate activities has been determined. Since the set Z determines any dependence between the choices for the applications of resources between aggregate activities, if we provide a systematic approach to modeling the flow types then we will have presented a systematic approach to modeling the dependence relationships.

To develop models of the flow types associated with the subnetworks to be analyzed in this Chapter, we first delimit the class of aggregate production networks that we will analyze

(Section 4.1). The definition of this class will allow us to define an *induced operating intensity* for each aggregate activity. The induced operating intensity will be a weighted average of the individual activity operating intensities and will have the interpretation of measuring the "progress" of an aggregate activity.

In Section 4.2, we present two classes of subnetworks for which we develop models of the flow types associated with the aggregate activities within the subnetwork. In a manner similar to how the flow types were modeled at the detailed level, the flow types for each aggregate activity will be determined by the induced operating intensity. By an analysis of the inventory balance constraints associated with the flow types at the detailed level, the flow types so modeled will be shown to be *consistent,* i.e., they satisfy the appropriate inventory balance constraints, and *reasonable,* i.e., the set Z induced from the models of the flow types contains collections of allocations to the aggregate activities which insure that there are schedules for the detailed activities consistent with the allocations which are on time and within budget.

The models developed are determined from the induced operating intensity. Hence, they are, in effect, dependent on the knowledge of the schedules for the underlying activities. To allow the models to be dependent on such knowledge would clearly defeat the purpose of aggregation. Thus, *as an absolute necessity,* the models constructed must be *independent* of any knowledge of the schedules for the underlying activities and therefore cannot be determined from the induced operating intensity.

In Section 4.3, we abstract from models of the flow types for the subnetworks analyzed in Section 4.2 to obtain models for flow types which are independent. This is done by modeling the domains of the induced operating intensities. Functions belonging to these domains are referred to as *aggregate operating intensities* and are not necessarily induced from a schedule for the underlying activities. Hence, the models of the flow types given in this section are "independent" from the underlying network. It will be shown that these models are "reasonable." The definitions of the models given in Section 4.2 serve to motivate the general definition for the aggregate network.

In Section 4.4, we introduce the technique of *replication* of detailed activities. This technique will enable us to identify classes of subnetworks which are, in fact, equivalent to subnetworks already analyzed. Two additional classes of subnetworks are analyzed through this technique.

In Section 4.5, we present some concluding remarks. We show how the models of the flow types can be approximated so that the constraints which define the set Z (of feasible allocations to the aggregate activities) are linear. We then point out how a Linear Program could be formulated to accomplish multi-project resource-use planning. Next, we give an example which shows that the models developed in Section 4.3 need not be "consistent." The example points out the need to further restrict the domain of the aggregate operating intensities. We also point out how the analyses carried out for the specific classes of subnetworks treated in this chapter can be extended for wider classes of networks. Finally, suggestions for future research are provided.

4.1. The Operating Intensity of an Aggregate Activity

In this section, we delimit the class of aggregate production networks G' that we will analyze. Within this class we can define an *operating intensity* for each aggregate activity. The operating intensity will be used to define the domains of the applications of exogenous resource flow types (the $y_l^{k'}s$). In a manner similar to Section 2.3.4, we will also use the operating intensity to define the domains of the other flow types.

Notation

To differentiate between activities at the detailed level from those at the aggregate level, we use lower case letters (possibly with subscripts) to denote detailed activities and upper case letters (possibly with subscripts) to denote aggregate activities. The symbol which denotes an aggregate activity will also be used to denote the set of detailed activities within the aggregate activity.

The class of aggregate production networks that we will analyze must first satisfy the following property.

Property I

For each aggregate activity A, there exist numbers α_{Al}, $l \in A$ such that if $\sum_{l \in A} b_{kl} > 0$ for some k then

$$\alpha_{Al} = \frac{b_{kl}}{\sum_{l \in A} b_{kl}}, \quad 1 \leq k \leq n. \tag{4.2}$$

Property I insists that each detailed activity within an aggregate activity must utilize the same percentage of the total amount of each resource required to complete all of the activities within the aggregate activity.

The Application of System Exogenous Resources

Let S denote a feasible schedule of start-time for the activities in G and let A denote an aggregate activity in G'. The application vector of system exogenous resources to A is easily seen to satisfy

$$y_A^k = \sum_{l \in A} y_l^k = \sum_{l \in A} b_{kl} z_l^S. \qquad (4.3)$$

(The last equality holds by (2.45). The definition of z_l^S is given by (2.56).)

Definition (4.4)

The *induced operating intensity for A derived from S*, denoted by z_A^S, is given by

$$z_A^S = \sum_{l \in A} \alpha_{Al} z_l^S.$$

If we let

$$a_{kA} = \sum_{l \in A} b_{kl} \qquad (4.5)$$

then by 4.3-4.5 it is simple to verify the identity

$$y_A^k = a_{kA} z_A^S. \qquad (4.6)$$

Any application vector of system exogenous inputs at the aggregate level satisfies the *form* given in (4.6). This is precisely the usual restriction imposed on the application vector of system exogenous inputs imposed by DLAAM (see Section 2.3.1). The interpretation of z_A^S is that the $\int_0^\tau z_A^S d\mu$ expresses the fraction of the total resources required to complete all of the detailed activities within A up to time τ. In effect, z_A^S is a way of measuring the *progress* of A towards completing the detailed activities within it.

Finally, in order to insure that the flow types at the aggregate level are not already determined it is necessary to insist that the class of aggregate production networks that we will analyze satisfies the following property.

Property II

For each aggregate activity A, $z_A^E \neq z_A^L$.

We insist on Property II because if z_A^E, the induced operating intensity derived from the early-start schedule, were equal to z_A^L, the induced operating intensity derived from the late-start schedule, then clearly all flow types at the aggregate level associated with A would be determined.

4.2. Developing Models of the Flow Types From the Induced Operating Intensity

In this section, models of the flow types associated with two classes of subnetworks, the Parallel $A:B$ Subnetwork and the Parallel $A:BC$ Subnetwork, are provided. The models of the flow types are shown to be consistent and reasonable as defined in Section 4.0. The models of the flow types will be determined by the induced operating intensity which also determines the application vector of system exogenous inputs to an aggregate activity.

4.2.1. The Parallel $A:B$ Subnetwork

The Parallel $A:B$ Subnetwork is shown in Figure (4-2). Here, detailed activities $a_1, a_2, \ldots, a_M$ are aggregated into aggregate A and detailed activities $b_1, b_2, \ldots, b_M$ are aggregated into aggregate B. In the aggregate subnetwork, there would be one arc from A to B.

In order to present our models, we first make two definitions.

Definition (4.7)

Let A denote an aggregate activity in G'. The *set of induced operating intensities for A derived from the set of feasible schedule of start-times for the detailed activities within A*, denoted by Z_A, is defined by

$$Z_A = \left\{ z_A \in L_+^\infty(\mu) \mid \exists \text{ a feasible schedule } S \text{ such that } z_A = z_A^S = \sum_{l \in A} \alpha_{Al} z_l^S \right\}.$$

Definition (4.8)

The *Parallel $A:B$ intermediate product transfer functional* is a map $f_{A:B}: Z_{a_1} \times \cdots \times Z_{a_M} \to Z_B$ defined by

$$f_{A:B}(z_{a_1}, \ldots, z_{a_M}) = \sum_{i=1}^{M} \alpha_{Bb_i} f_{a_i b_i}(z_{a_i}).$$

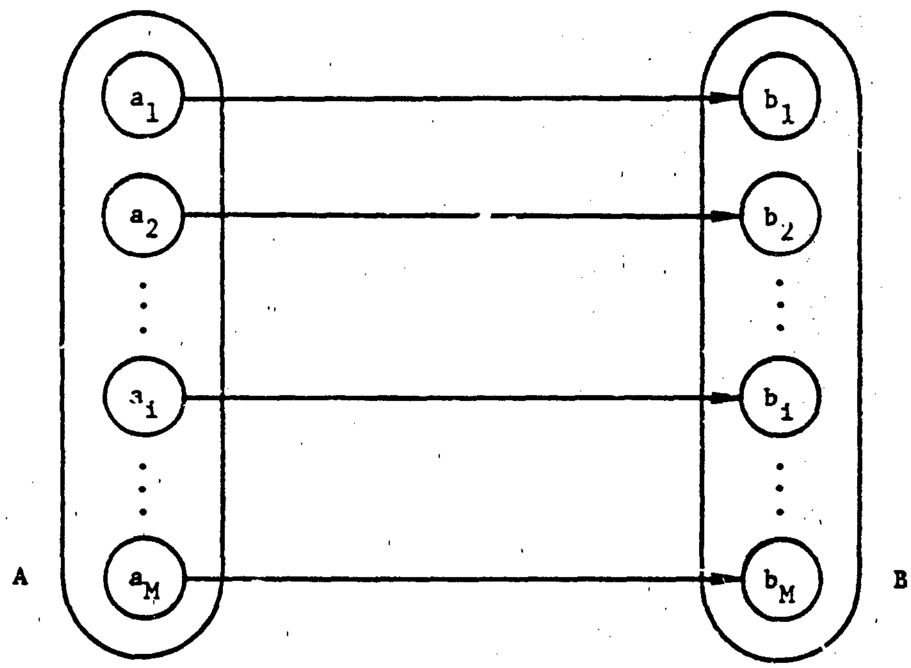


FIGURE (4-2)

THE PARALLEL A:B SUBNETWORK

See (2.45) in Section (2.3.4) for the definition of the function $f_{a_i b_i}(z_{a_i})$.

The models of the flow types associated with the Parallel $A:B$ Subnetwork are determined by the induced operating intensity in the following manner. If, for $1 \leqslant k \leqslant n$,

$$y_A^k = a_{kA} z_A \quad \text{where} \quad z_A = \sum_{i=1}^{M} \alpha_{Aa_i} z_{a_i}$$

$$y_B^k = a_{kB} z_B \quad \text{where} \quad z_B = \sum_{i=1}^{M} \alpha_{Bb_i} z_{b_i}$$

then the flow types $F_A(y_A, W_A)$, V_{AB}, and W_B associated with the Parallel $A:B$ Subnetwork are modeled by

$$F_A(y_A, W_A) = \sum_{i=1}^{M} \alpha_{Bb_i} F_{a_i}(y_{a_i}, W_{a_i}) \tag{4.9}$$

$$= \sum_{i=1}^{M} \alpha_{Bb_i} z_{a_i} \qquad \text{(by 2.46)}$$

$$V_{AB} = f_{A:B}(z_{a_1}, \ldots, z_{a_M}) \tag{4.10}$$

$$= \sum_{i=1}^{M} \alpha_{Bb_i} f_{a_i b_i} \qquad \text{(by 4.8)}$$

$$= \sum_{i=1}^{M} \alpha_{Bb_i} V_{a_i b_i} \qquad \text{(by 2.51)}$$

$$W_B = \sum_{i=1}^{M} \alpha_{Bb_i} W_{b_i} \tag{4.11}$$

$$= \sum_{i=1}^{M} \alpha_{Bb_i} z_{b_i} \qquad \text{(by 4.4)}$$

$$= z_B .$$

It is clear by definitions 4.9-4.11 that we are assuming that only *one* intermediate product is being "sent" to B. To assume that M products were being produced by A and sent to B would defeat the whole point of aggregation. One would simply define, for $k = 1,2,\ldots,M$,

$$F_A^k(y_A, W_A) = F_{a_k}(y_{a_k}, W_{a_k})$$

$$V_{AB}^k = f_{a_k b_k}(z_{c_k})$$

$$W_B^k = z_{b_k}$$

which is just another way of describing the flow types at the detailed level.

We now argue that the models of the flow types defined in 4.9-4.11 are consistent and reasonable. For the Parallel $A:B$ Subnetwork, there is only one inventory balance constraint relevant to A which must be satisfied,

$$\int_0^\tau \{F_A(y_A, W_A) - V_{AB}\} d\mu \geqslant 0, \quad \forall \tau \in R_+.[1] \tag{4.12}$$

Fix $z_{a_i} \in Z_{a_i}$, $i = 1,2,\ldots,M$. Substituting in (4.12) the definitions for $F_A(y_A, W_A)$ and V_{AB} we obtain that, $\forall \tau \in R_+$,

$$\int_0^\tau \{F_A(y_A, W_A) - V_{AB}\} d\mu = \int_0^\tau \left\{ \sum_{i=1}^M \alpha_{Bb_i} F_{a_i}(y_{a_i}, W_{a_i}) - \sum_{i=1}^M \alpha_{Bb_i} V_{a_i b_i} \right\} d\mu$$

$$= \sum_{i=1}^M \alpha_{Bb_i} \int_0^\tau \{F_{a_i}(y_{a_i}, W_{a_i}) - V_{a_i b_i}\} d\mu$$

$$\geqslant 0.$$

[1] The set on which are integrating includes the endpoints 0 and τ.

Thus, we have shown that the models of the flow types $F_A(y_A, W_A)$ and V_{AB} are consistent.

For the Parallel $A{:}B$ Subnetwork, there is only one inventory balance constraint relevant to *both* A and B. This is the constraint

$$\int_0^{\tau} \{V_{AB} - W_B\} d\mu \geqslant 0, \quad \forall \tau \in R_+. \tag{4.13}$$

Substituting in the definitions for V_{AB} and W_B into (4.13) we obtain that, $\forall \tau \in R_+$,

$$\int_0^{\tau} \{V_{AB} - W_B\} d\mu = \int_0^{\tau} \left\{ \sum_{i=1}^{M} \alpha_{Bb_i} V_{a_i b_i} - \sum_{i=1}^{M} \alpha_{Bb_i} z_{b_i} \right\} d\mu$$

$$= \sum_{i=1}^{M} \int_0^{\tau} \{V_{a_i b_i} - z_{b_i}\} d\mu$$

$$\geqslant 0. \tag{4.14}$$

Constraint (4.14) imposes a restriction on the choices for the induced operating intensities for A and B. Since the choices for the induced operating intensities for A and B determine the applications of system exogenous inputs to A and B, *constraint (4.14) is the model of the dependence relationship between A and B for the Parallel $A{:}B$ Subnetwork.* Thus, the question of reasonableness of the models of the flow types for the Parallel $A{:}B$ Subnetwork reduces to the question of how reasonable is (4.14) as a model of the dependence relationship between A and B.

At the detailed level, M inventory balance constraints

$$\int_0^{\tau} \{V_{a_i b_i} - z_{b_i}\} d\mu \geqslant 0, \quad \forall \tau \in R_+, \; i = 1, 2, \ldots, M$$

insure that the start-times for b_i, $i=1,2,\ldots,M$, are consistent with the finish times for a_i, $i=1,2,\ldots,M$. At the aggregate end, only *one* inventory balance constraint (4.14) exists to constrain the possible start-times for b_i, $i=1,2,\ldots,M$, given the finish times for a_i, $i=1,2,\ldots,M$ (as reflected by z_A). Hence, with only one constraint it will not be possible to model the dependence relationships *exactly*. But, for our models of the flow types, it is immediate by (4.14) that, for a fixed z_A, all z_B's whose start-times for the b_i's are consistent with z_A *do* satisfy (4.14) Thus, our models of the flow types are "reasonable" in the sense described in Section 4.0.

We make two comments about the subnetwork just analyzed. First, this one example illustrates that modeling the so-called dependence relationships between the applications of system exogenous inputs to aggregate activities is encompassed by modeling the flow types associated with the intermediate product transfers between aggregate activities. Second, for the subnetwork just analyzed, the production function $F_A(y_A, W_A)$ was *not* equal to the induced operating intensity z_A. This differs from the detailed case where $F_{a_i}(y_{a_i}, W_{a_i}) = z_{a_i}$, $i=1,2,\ldots,M$.

The reason why we cannot model $F_A(y_A, W_A)$ as z_A is because z_A is a measurement of the rate A utilizes its resources, *not* a measurement of when the activities within A have been completed. To illustrate this point, consider the following example of a Parallel $A:B$ Subnetwork with 2 activities:

(a) the duration of each of the 4 activities is one period,

(b) $\alpha_{a_1} = \alpha_{b_2} = .05$, $\alpha_{a_2} = \alpha_{b_1} = .95$,

(c) $z_{a_1} = 1_{(0,1)}$, $z_{a_2} = 1_{(2,3)}$.

Suppose $F_A(y_A, W_A)$ were modeled as z_A. In this example at $\tau = 2$, z_A tells us that A has completed 5% of its work. As measured by *resource use*, this is true. However, at $\tau = 2$, a_1 has finished and we would desire to *allow* b_1 to start. If b_1 did start at $\tau = 2$, then the following problem would emerge:

$$\int_0^2 F_A(y_A, W_A)\,d\mu = \int_0^2 z_A\,d\mu = \int_0^2 \{\alpha_{Aa_1} z_{a_1} + \alpha_{Aa_2} z_{a_2}\}\,d\mu$$

$$= .05$$

$$\geqslant \int_0^2 V_{AB}\,d\mu \geqslant \int_0^2 z_B\,d\mu - \int_0^2 \{\alpha_{Bb_1}z_{b_1} + \alpha_{Bb_2}z_{b_2}\}\,d\mu$$

$$= .95.$$

So, for *no* choice of V_{AB} would it be possible to begin b_1 at time $\tau = 2$. This eliminates a possibility for b_1 which is not desirable.

4.2.2. The Parallel A:BC Subnetwork

The *Parallel A:BC Subnetwork* is shown in Figure (4-3). Here, detailed activities $a_1, a_2, \ldots, a_K, a_{K+1}, \ldots, a_M$ were aggregated into A, detailed activities $b_1, b_2, \ldots, b_K$ were aggregated into aggregate B, and $c_1, c_2, \ldots, c_L$ were aggregated into aggregate C.

The flow types associated with the Parallel A:BC Subnetwork are $F_A(y_A, W_A)$, V_{AB}, V_{AC}, W_B, and W_C. The inventory balance constraints associated with these flow types are:

$$\int_0^\tau \{F_A(y_A, W_A) - (V_{AB} + V_{AC})\}\,d\mu \geqslant 0, \quad \forall \tau \in R_+, \tag{4.15}$$

$$\int_0^\tau \{V_{AB} - W_B\}\,d\mu \geqslant 0, \quad \forall \tau \in R_+, \tag{4.16}$$

$$\int \{V_{AC} - W_C\}\,d\mu \geqslant 0, \quad \forall \tau \in R_+. \tag{4.17}$$

Our analysis of the Parallel A:B Subnetwork motivates the following *structure* for the models of the flow types:

For some $d_1 \in R_+$, $d_2 \in R_+$, $d_1 + d_2 = 1$, if, for $1 \leqslant k \leqslant n$,

$$y_A^k = a_{kA} z_A \quad \text{for } z_A = \sum_{i=1}^{M} \alpha_{Aa_i} z_{a_i}$$

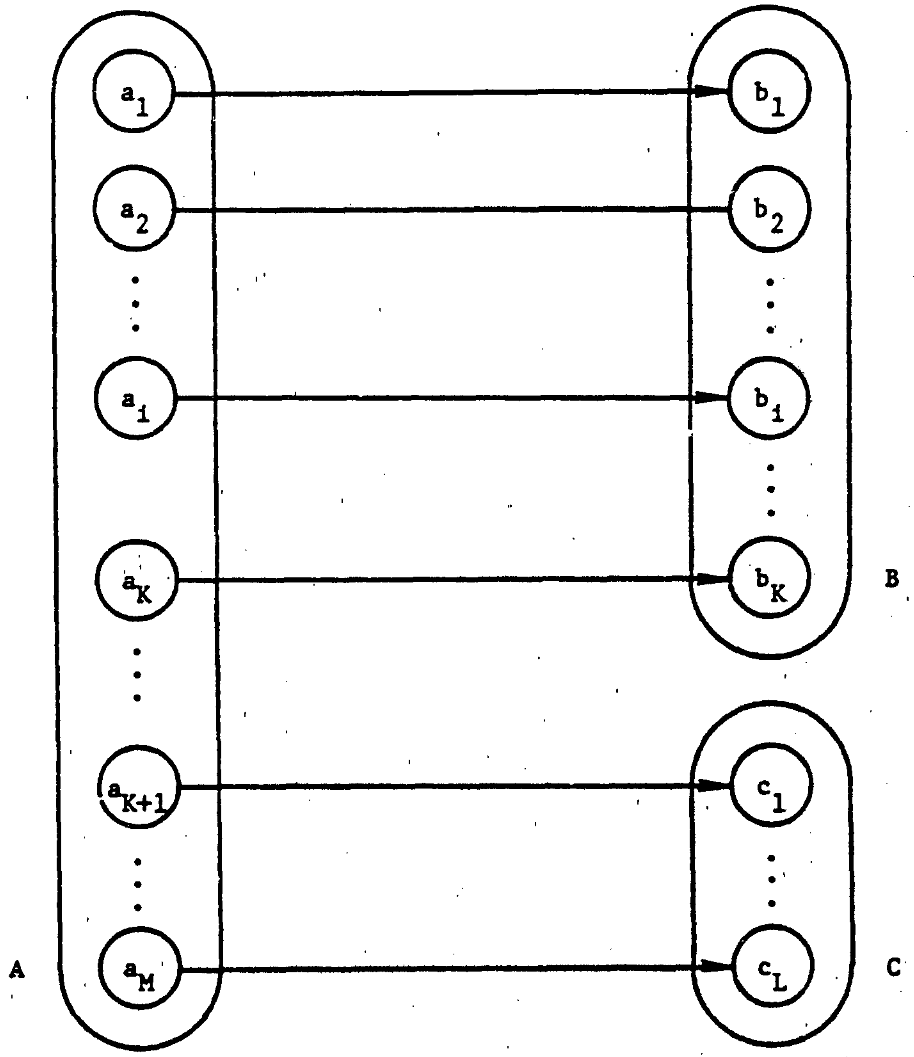


FIGURE (4-3)

THE PARAL EL A:BC SUBNETWORK

$$y_B^k = a_{kB} z_B \quad \text{for } z_B = \sum_{i=1}^{K} \alpha_{Bb_i} z_{b_i}$$

$$y_C^k = a_{kC} z_C \quad \text{for } z_C = \sum_{i=1}^{L} \alpha_{Cc_i} z_{c_i}$$

then

$$F_A(y_A, W_A) = d_1\left\{\sum_{i=1}^{K} \alpha_{Bb_i} z_{a_i}\right\} + d_2\left\{\sum_{i=1}^{L} \alpha_{Cc_i} z_{a_{i+K}}\right\} \tag{4.18}$$

$$V_{AB} = d_1\left\{\sum_{i=1}^{k} \alpha_{Bb_i} f_{a_i b_i}(z_{a_i})\right\} \tag{4.19}$$

$$V_{AC} = d_2\left\{\sum_{i=1}^{L} \alpha_{Cc_i} f_{a_{i+K} b_i}(z_{a_{i+K}})\right\} \tag{4.20}$$

$$W_B = d_1 z_B \tag{4.21}$$

$$W_C = d_2 z_C \,. \tag{4.22}$$

It is easy to verify that the models of the flow types defined by (4.18-4.22) satisfy (4.15-4.17) and thus are consistent.

Let A_1 denote the aggregate activity comprised by $a_1, a_2, \ldots, a_K$ and let A_2 denote the aggregate activity comprised of $a_{K+1}, \ldots, a_M$. If $F_{A_j}(y_{A_j}, W_{A_j})$, $j = 1,2$, is defined by (4.9) and $f_{A_1 B}$, $f_{A_2 C}$ are defined by (4.8) then (4.18-4.20) become

$$F_A(y_A, W_A) = d_1 F_{A_1}(y_{A_1}, W_{A_1}) + d_2 F_{A_2}(y_{A_2}, W_{A_2}) \tag{4.23}$$

$$V_{AB} = d_1 f_{A_1:B}(z_{a_1}, \ldots, z_{a_K}) = d_1 V_{A_1:B} \tag{4.24}$$

$$V_{AC} = d_2 f_{A_2:C}(z_{a_{K+1}}, \ldots, z_{a_M}) = d_2 V_{A_2 C} \tag{4.25}$$

What 4.23-4.25 show is that the Parallel $A:BC$ Subnetwork ($K + L = M$) is, in effect, a combination of a Parallel $A_1:B$ Subnetwork and an Parallel $A_2:C$ Subnetwork. What 4.23-4.25 also

shows is that it is not possible to arrive at independent models for the flow types V_{AB} and V_{AC}. This is because V_{AB} (V_{AC}) is determined only from the z_{A_1} (z_{A_2}) part of z_A. Hence, we will model this sum $V_A = V_{AB} + V_{AC}$. It will be convenient for later purposes to introduce the following definition.

Definition (4.26)

The Parallel $A:BC$ intermediate product functional is a map

$$f_{A:BC}: Z_{a_1} \times \cdots \times Z_{a_M} \to d_1 Z_B + d_2 Z_C = \left\{ d_1 z_B + d_2 z_C \mid z_B \in Z_B ,\ z_C \in Z_C \right\}$$

defined by

$$f_{A:BC}(z_{a_1}, \ldots, z_{a_M}) = d_1 f_{A_1:B}(z_{c_1}, \ldots, z_{a_K}) + d_2 f_{A_2:C}(z_{a_{K+1}}, \ldots, z_{a_M}).$$

Of course, $V_A = f_{A:BC}(z_{a_1}, \ldots, z_{a_M})$.

The inventory balance constraint associated with V_A, W_B, and W_C is

$$0 \le \int_0^{\tau} \{V_A - (W_B + W_C)\} d\mu , \quad \forall \tau \in R_+$$

which reduces to

$$0 \le \int_0^{\tau} \left\{ d_1 (f_{A_1:B}(z_{a_1}, \ldots, z_{a_K}) - z_B) + d_2 (f_{A_2:C}(z_{a_{K+1}}, \ldots, z_{a_M}) - z_C) \right\} d\mu . \qquad (4.27)$$

for $z_A \in Z_A$. It is immediate by (4.27) that all start-times for $b_1, \ldots, b_K$, $c_1, \ldots, c_L$ which are consistent with the finish times for $a_1, \ldots, a_M$ satisfy (4.27). Hence, our models of the flow types are "reasonable" in the sense described in Section 4.0. The appropriate choice for d_1 and d_2 is deferred until the next section.

4.3. Constructing Independent Models of the Flow Types

The models of the flow types in the previous section were determined by the induced operating intensity and hence were not independent. To construct independent models, we first model the domain of the induced operating intensity. Functions belonging to this domain will be referred to as *aggregate operating intensities*. The models of the flow types provided in this section are determined from the aggregate operating intensities.

4.3.1. Constructing the Domain of the Aggregate Operating Intensity

Let A represent an aggregate activity. As notation, let

$$E_A = \min_{l \in A} E_l, \quad L_A = \max_{l \in A} L_l, \quad W_A = [E_A, L_A]. \tag{4.28}$$

(E_l is the early-start time for activity l and L_l is the late-start time for activity l.) It is immediate by the definition of the induced operating intensity (4.7) that if $z_A \in Z_A$ then z_A would satisfy the following boundary conditions:

$$z_A \textit{ is a step-function} \tag{4.29}$$

$$\int_{E_A}^{L_A} z_A \, d\mu = 1 \tag{4.30}$$

$$\int_{E_A}^{\tau} z_A^L \, d\mu \leq \int_{E_A}^{\tau} z_A \, d\mu \leq \int_{E_A}^{\tau} z_A^E \, d\mu, \quad \forall \tau \in W_A. \tag{4.31}$$

Let D_A denote the set of all functions in $L_+^\infty(\mu)$ satisfying 4.29-4.31. It is immediate by the definition of Z_A that $D_A \supset Z_A$. D_A will be taken as the model of Z_A. Functions belonging to D_A are referred to as ***aggregate operating intensities*** for aggregate A. Clearly, the functions belonging to D_A are independent (in the sense described in Section 4.0).

4.3.2. Constructing an Independent Model of the Application Vector of System Exogenous Inputs

In Section 4.2, y_A, the application vector of system exogenous inputs to aggregate A, was determined by an induced operating intensity for aggregate A, $z_A \in Z_A$. We will model the application vector y_A in a similar manner. That is, if y_A denotes an application vector of system exogenous inputs to aggregate A then it is assumed that, for $1 \leqslant k \leqslant n$,

$$y_A^k = a_{kA} z_A \quad \text{for some } z_A \in D_A. \tag{4.32}$$

4.3.3. Constructing an Independent Model of V_{AB} for the Parallel $A{:}B$ Subnetwork

For the Parallel $A{:}B$ Subnetwork, the intermediate product transfer variable V_{AB} was determined from the induced operating intensity via the intermediate product transfer functional

$$f_{A:B}: Z_{a_1} \times \cdots \times Z_{a_M} \to Z_B.$$

(see (4.8) for the definition of $f_{A:B}$). In this section, we will construct an independent model for V_{AB} which we will denote by V_{AB}^*. The essential idea is to construct an $f_{A:B}^*: D_A \to D_B$ which reasonably approximates $f_{A:B}$ on $Z_{a_1} \times \cdots \times Z_{a_M}$ and then define $V_{AB}^* = f_{A:B}^*(z_A)$ for $z_A \in D_A$.

An inspection of the definition for the intermediate product transfer functional $f_{A:B}$ shows that it satisfies the following boundary conditions, $\forall (z_{a_1}, \ldots, z_{a_M}) \in Z_{a_1} \times \cdots \times Z_{a_M}$, and $\forall \tau \in W_B$:

$$\int_{E_B}^{L_B} f_{A:B}(z_{a_1}, \ldots, z_{a_M})\, d\mu = 1 \tag{4.33}$$

$$\int_{E_B}^{\tau} f_{A:B}(z_{a_1}^L, \ldots, z_{a_M}^L)\, d\mu \leqslant \int_{E_B}^{\tau} f_{A:B}(z_{a_1}, \ldots, z_{a_M})\, d\mu \leqslant \int_{E_B}^{\tau} f_{A:B}(z_{a_1}^E, \ldots, z_{a_M}^E)\, d\mu \tag{4.34}$$

It is also easy to see that

$$f_{A:B}(z^L_{a_1}, \ldots, z^L_{a_M}) = z^L_B \tag{4.35}$$

$$f_{A:B}(z^E_{a_1}, \ldots, z^E_{a_M}) = z^E_B . \tag{4.36}$$

Our goal is to construct an $f^*_{A:B}$ on D_A which reasonably approximates $f_{A:B}$ on $Z_{a_1} \times \cdots \times Z_{a_M}$. In view of the definition of D_A (4.29-4.31) and the boundary conditions 4.33-4.36, it seems reasonable to insist that $f^*_{A:B}$ satisfies the following boundary conditions, $\forall$ $z_A \in D_A$, and $\forall \tau \in W_B$:

$$\int_{E_B}^{L_B} f^*_{A:B}(z_A)\,d\mu = 1 \tag{4.37}$$

$$\int_{E_B}^{\tau} f_{A:B}(z^L_{a_1}, \ldots, z^L_{a_M})\,d\mu \leq \int_{E_B}^{\tau} f^*_{A:B}(z_A)\,d\mu \leq \int_{E_B}^{\tau} f_{A:B}(z^E_{a_1}, \ldots, z^E_{a_M})\,d\mu \tag{4.38}$$

which by 4.35-4.36 is equivalent to

$$\int_{E_B}^{\tau} z^L_B\,d\mu \leq \int_{E_B}^{\tau} f^*_{A:B}(z_B)\,d\mu \leq \int_{E_B}^{\tau} z^E_B\,d\mu$$

$$f^*_{A:B}(z^E_A) = f_{A:B}(z^E_A) = z^E_B \tag{4.39}$$

$$f^*_{A:B}(z^L_A) = f_{A:B}(z^L_A) = z^L_B .$$

Fix $z_A \in D_A$. The only real information that we know about z_A is given in the boundary conditions 4.29-4.31 and this is not much. Since $f^*_{A:B}(z_A)$ satisfies 4.37-4.39, an intuitively appealing idea for the definition of $f^*_{A:B}(z_A)$ is that $f^*_{A:B}(z_A)$ should satisfy the following property:

For some $\rho: W_B \to W_A$, $f^*_{A:B}(z_A)$ must satisfy the following equation:

$$\frac{\int_{E_B}^{\tau} \{f^*_{A:B}(z_A) - z_B^L\} d\mu}{\int_{E_B}^{\tau} \{z_B^E - z_B^L\} d\mu} = \frac{\int_{E_A}^{\rho(\tau)} \{z_A - z_A^L\} d\mu}{\int_{E_A}^{\rho(\tau)} \{z_A^E - z_A^L\} d\mu}, \quad \forall \tau \in W_B. \tag{4.40}$$

A map ρ is required since $W_B \neq W_A$. Of course, to transform the idea of (4.40) into a well-defined mathematically correct definition for $f^*_{A:B}(z_A)$ requires us to impose certain restrictions on ρ and the right-hand side in (4.40)

To arrive at the restrictions, it will be convenient to introduce the following definition.[1]

Definition (4.41)

The *relative progress functional* for an aggregate activity A is a map $p_A: D_A \times R_+ \to [0,1]$ defined by

$$p_A(z_A, t) = \begin{cases} 0 & \text{if } t \notin W_A \\ \dfrac{\int_{E_A}^{t} \{z_A - z_A^L\} d\mu}{\int_{E_A}^{t} \{z_A^E - z_A^L\} d\mu} & \text{if } t \in W_A,\ z_A \neq z_A^E \\ 1 & \text{if } t \in W_A,\ z_A = z_A^E \end{cases}$$

$p_A(z_A, t)$ is referred to as "the relative progress" of z_A at time t.

Let us re-write (4.40) as

$$\int_{E_B}^{\tau} f^*_{A:B}(z_A) d\mu = \int_{E_B}^{\tau} z_B^L d\mu + p_A(z_A, \rho(\tau)) \int_{E_B}^{\tau} \{z_B^E - z_B^L\} d\mu, \quad \forall \tau \in W_B. \tag{4.42}$$

[1] First introduced in Leachman, Boysen [1982].

Equation (4.42) is an implicit definition for $f^*_{A:B}(z_A)$. Since indefinite integrals of integrable functions are absolutely continuous and hence differentiable[2], if ρ were differentiable then one could differentiate both sides of (4.42) with respect to τ to obtain a definition for $f^*_{A:B}(z_A)$. The criterion of differentiability on ρ is not too restrictive. Equation (4.42) and the definition for $p_A(z_A, \rho(\tau))$ suggest that it is reasonable to insist that ρ should be

$$\text{continuous and increasing} \tag{4.43}$$

and satisfy the boundary conditions

$$\rho(E_B) = E_A\,, \;\; \rho(L_B) = L_A \tag{4.44}$$

If so, then ρ would be automatically differentiable.[3] Hence, to complete the definition for $f^*_{A:B}(z_A)$ as given in (4.42) and hence V^*_{AB} it is sufficient to select a ρ which satisfies 4.43-4.44.

A natural choice for ρ which does satisfy 4.43-4.44 is to insist that

$$\frac{\rho(\tau)-E_A}{L_A-E_A} = \frac{\tau-E_B}{L_B-E_B}\,, \quad \forall \tau \in W_B$$

which re-written becomes

$$\rho(\tau) = E_A + \frac{(L_A-E_A)}{(L_B-E_B)}(\tau-E_B)\,, \quad \forall \tau \in W_B\,. \tag{4.45}$$

However, this will not be a good choice for ρ. To motivate why, consider the special case when $L_A - E_A = L_B - E_B$. This would occur, for example, if the detailed activities within A and B had equal durations. Under this special case, (4.45) reduces to the requirement that

$$\rho(\tau) = E_A + (\tau - E_B) = \tau - (E_B - E_A)\,, \quad \forall \tau \in W_B\,.$$

[2] almost everywhere.

[3] See Royden p. 96.

Hence, ρ would be a simple time lag of length $E_B - E_A$. Since the definition for ρ as suggested in (4.45) does not in any way consider the boundaries z_A^E, z_A^L, z_B^E, and z_B^L, it follows that under the special case cited above, ρ would always be a time lag length $E_B - E_A$ *regardless* of the boundaries. In general, the boundaries z_B^E and z_B^L are *not* determined from z_A^E and z_A^L by a simple time lag. Thus, (4.45) should not be a good choice for ρ.

To illustrate further why the choice for ρ as given by (4.45) is not a good one and to motivate our choice for ρ, consider the example of a 2-Parallel $A:B$ Subnetwork shown in Figure (4-4). The dotted line in Figure (4-4a) represents the curve $\int_{E_B}^{x} z_A \, d\mu$ generated from a particular $z_A \in D_A$. The dotted line in Figure (4-4b) represents the curve $\int_{E_B}^{x} f_{A:B}^{*}(z_A) \, d\mu$ as defined by (4.42) with ρ defined by (4.45).

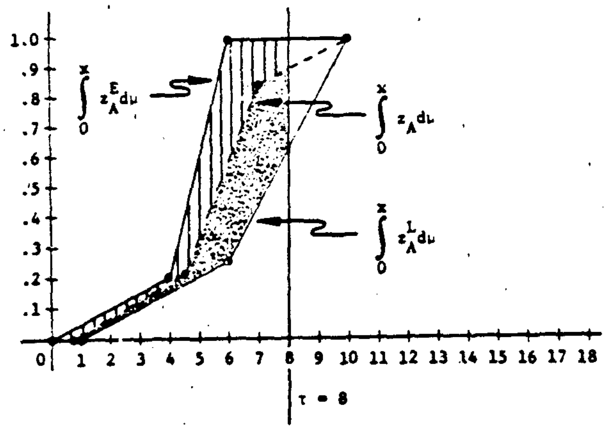
When we modeled V_{AB} in Section 4.2.1, we motivated that it is through the inventory balance constraint

$$\int_{E_B}^{\tau} (V_{AB} - z_B) \, d\mu \geqslant 0, \quad \forall \tau \in W_B$$

that the dependence relationship between the applications of system exogenous inputs to A and B is modeled. Since we have modeled V_{AB} by V_{AB}^{*} it is now through the inventory balance constraint

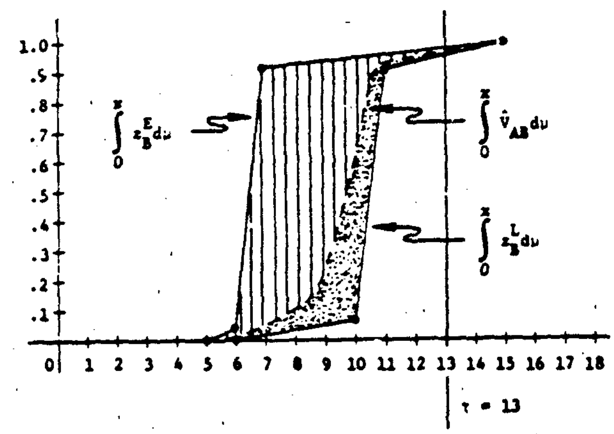
$$\int_{E_B}^{\tau} (f_{A:B}^{*}(z_A) - z_B) \, d\mu \geqslant 0, \quad \forall \tau \in W_B \tag{4.46}$$

that the dependence relationship between A and B is modeled. Hence, for each $\tau \in W_B$, the restriction imposed on the choice for z_B given a choice for z_A may be measured by the shaded area shown in Figure (4-4b). In our example, for any $\tau \in W_B$, the shaded area evaluated at τ in proportion to the total area evaluated at τ is quite small. This reflects the fact that by our choice of $f_{A:B}^{*}(z_A)$, i.e., V_{AB}^{*} constraint (4.46) is forcing z_B to essentially "run late." On the



	a_1	a_2
Duration	5	2
Early-start	0	4
Late-start	1	5
Resource use	$\alpha_{Aa_1} = .25$	$\alpha_{Aa_2} = .75$

(a)



	b_1	b_2
Duration	9	1
Early-start	5	6
Late-start	6	10
Resource use	$\alpha_{Bb_1} = .10$	$\alpha_{Bb_2} = .90$

(b)

FIGURE (4-4)

EXAMPLE OF INTERMEDIATE PRODUCT TRANSFER CURVE
DERIVED FROM A SIMPLE TIME LAG

other hand, an inspection of Figure (4-4a) reveals that the shaded area evaluated at $\rho(\tau)$ in proportion to the total area evaluated at $\rho(\tau)$ is fairly large. This reflects the fact that z_A is essentially "running early." So what we have is that z_A is "running early" but we are constraining z_B to run "late".

Based on this one example, let us insist that regardless of the choice for ρ, $f^*_{A:B}(z_A)$ must satisfy the additional property that the shaded areas normalized by the total areas, evaluated at the corresponding points in time, are equal. Mathematically, we are insisting that $f^*_{A:B}(z_A)$ satisfies the following criterion:

$$\frac{\int\limits_{E_B}^{\tau}\int\limits_{E_B}^{x}[f^*_{A:B}-z^L_B]\,d\mu\,dx}{\int\limits_{E_B}^{\tau}\int\limits_{E_B}^{x}[z^E_B-z^L_B]\,d\mu\,dx}=\frac{\int\limits_{E_A}^{\rho(\tau)}\int\limits_{E_A}^{x}[z_A-z^E_A]\,d\mu\,dx}{\int\limits_{E_A}^{\rho(\tau)}\int\limits_{E_A}^{x}[z^E_A-z^L_A]\,d\mu\,dx},\quad \forall\tau\in W_B. \tag{4.47}$$

It will be convenient to introduce the following notation: for an aggregate activity A,

$$h_A(x)=\int\limits_{0}^{x}[z^E_A-z^L_A]\,d\mu,\qquad x\in R_+ \tag{4.48}$$

(Note that $h_A(x)=h_A(L_A)$, $\forall x\geqslant L_A$.)

Re-write (4.47), $\forall\tau\in W_B$, as

$$\int\limits_{E_B}^{\tau}\int\limits_{E_B}^{x}f^*_{A:B}(z_A)\,d\mu\,dx=\int\limits_{E_B}^{\tau}\int\limits_{E_B}^{x}z^L_B\,d\mu\,dx+\frac{\int\limits_{E_B}^{\tau}h_B(x)\,dx\int\limits_{E_A}^{\rho(\tau)}p_A(z_A,x)h_A(x)\,dx}{\int\limits_{E_A}^{\rho(\tau)}h_A(x)\,dx}. \tag{4.49}$$

Proposition (4.3.1)

If $\forall z_A \in D_A$, $f^*_{A:B}(z_A)$ is required to satisfy (4.49) for some ρ satisfying 4.43-4.44 then $f^*_{A:B}(z_A)$ satisfies (4.42) $\forall z_A \in D_A$ for the same ρ if and only if ρ is the (unique) solution to the equation

$$\frac{\int_{E_B}^{\tau} h_B(x)\,dx}{h_B(L_B)} = \frac{\int_{E_A}^{\rho(\tau)} h_A(x)\,dx}{h_A(L_A)}, \quad \forall \tau \in W_B. \tag{4.50}$$

Proof of Proposition (4.3.1)

Suppose $\forall z_A \in D_A$ and for some ρ satisfying 4.43-4.44, $f^*_{A:B}(z_A)$ satisfies *both* (4.42) and (4.49). Differentiate each side of the equation (4.49) with respect to τ to obtain the equation, $\forall \tau \in W_B, \forall z_A \in D_A$,

$$\int_{E_B}^{\tau} f^*_{A:B}(z_A)\,d\mu = \int_{E_B}^{\tau} z_B^1\,d\mu + p_A(z_A,(\rho(\tau))h_A(\rho(\tau))\rho'(\tau)\frac{\int_{E_B}^{\tau} h_B(x)\,dx}{\int_{E_A}^{\rho(\tau)} h_A(x)\,dx}$$

$$+\int_{E_A}^{\rho(\tau)} p_A(z_A,x)h_A(x)\,dx \; \frac{h_B(\tau)\int_{E_A}^{\rho(\tau)} h_A(x)\,dx - h_A(\rho(\tau))\rho'(\tau)\int_{E_B}^{\tau} h_B(x)\,dx}{\left[\int_{E_A}^{\rho(\tau)} h_A(x)\,dx\right]^2}. \tag{4.51}$$

Since $f^*_{A:B}(z_A)$ also satisfies (4.42), if we subtract (4.42) from (4.51) and rearrange terms we obtain the equation, $\forall \tau \in W_B, \forall z_A \in D_A$,

$$0 = \left[h_B(\tau) - h_A(\rho(\tau))\rho'(\tau)\frac{\int_{E_B}^{\tau} h_B(x)\,dx}{\int_{E_A}^{\rho(\tau)} h_A(x)\,dx}\right]\left[p_A(z_A,(\rho(\tau)) - \frac{\int_{E_A}^{\rho(\tau)} p_A(z_A,x)h_A(x)\,dx}{\int_{E_A}^{\rho(\tau)} h_A(x)\,dx}\right]. \tag{4.52}$$

Since $h_A(x) > 0$ (except at $x = E_A$) then it is clear that

$$p_A(z_A, \rho(\tau)) - \frac{\int_{E_A}^{\rho(\tau)} p_A(z_A, x) h_A(x) dx}{\int_{E_A}^{\rho(\tau)} h_A(x) dx} \neq 0, \quad \forall \tau \in W_B \tag{4.53}$$

$p_A(z_A, x)$ is strictly increasing. Clearly, D_A contains z_A's such that $p_A(z_A, x)$ is strictly increasing (by Property I, Section 4.1). Since (4.52) holds $\forall z_A \in D_A$ and since the expression

$$h_B(\tau) - h_A(\rho(\tau))\rho'(\tau) \frac{\int_{E_B}^{\tau} h_B(x) dx}{\int_{E_A}^{\rho(\tau)} h_A(x) dx}$$

is independent of z_A, it follows by (4.53) that

$$0 = h_B(\tau) - h_A(\rho(\tau))\rho'(\tau) \frac{\int_{E_B}^{\tau} h_B(x) dx}{\int_{E_A}^{\rho(\tau)} h_A(x) dx}$$

or, equivalently,

$$\frac{h_B(\tau)}{\int_{E_B}^{\tau} h_B(x) dx} = \frac{h_A(\rho(\tau))\rho'(\tau)}{\int_{E_A}^{\rho(\tau)} h_A(x) dx}, \quad \forall \tau \in W_B . \tag{4.54}$$

Since

$$\frac{h_B(\tau)}{\int_{E_B}^{\tau} h_B(x) dx} = \frac{d}{d\tau} \left\{ \ln \int_{E_B}^{\tau} h_B(x) dx \right\}, \quad \forall \tau \in W_B$$

$$\frac{h_A(\rho(\tau))\rho'(\tau)}{\int\limits_{E_A}^{\rho(\tau)} h_A(x)dx} = \frac{d}{d\tau}\left\{\ln \int\limits_{E_A}^{\rho(\tau)} h_A(x)dx\right\}, \quad \forall \tau \in W_B$$

then by integrating both sides of (4.54) we obtain the equation

$$\int\limits_{E_B}^{\tau} h_B(x)dx = c\cdot \int\limits_{E_A}^{\rho(\tau)} h_A(x)dx\,, \quad \forall \tau \in W_B\,, \tag{4.55}$$

for some constant c. Evaluating (4.55) for $\tau = L_B$ shows that the constant c equals $\frac{h_B(L_B)}{h_A(L_A)}$.

Thus, ρ must satisfy the equation

$$\frac{\int\limits_{E_B}^{\tau} h_B(x)dx}{h_B(L_B)} = \frac{\int\limits_{E_A}^{\rho(\tau)} h_A(x)dx}{h_A(L_A)}, \quad \forall \tau \in W_B$$

which proves the desired result.■

To prove the converse direction, suppose ρ satisfies (4.50). Clearly, ρ is continuous, increasing, unique and satisfies $\rho(E_B) = E_A$, $\rho(L_B) = L_A$. Since ρ satisfies (4.50), (4.49) becomes, $\forall \tau \in W_B$,

$$\int\limits_{E_B}^{\tau}\int\limits_{E_B}^{x} f^{*}_{A:B}(z_A)d\mu\, dx = \int\limits_{E_B}^{\tau}\int\limits_{E_B}^{x} z_B^{*} d\mu\, dx + \frac{h_B(L_B)}{h_A(L_A)} \frac{\int\limits_{E_A}^{\rho(\tau)} p_A(z_A, x)h_A(x)dx}{\int\limits_{E_A}^{\rho(\tau)} h_A(x)dx}. \tag{4.56}$$

Differentiating each side of (4.56) with respect to τ we obtain the equation, $\forall \tau \in W_B$,

$$\int\limits_{E_B}^{\tau} f^{*}_{A:B}(z_A)d\mu = \int\limits_{E_B}^{\tau} z_B^{*} d\mu + \frac{h_B(L_B)}{h_A(L_A)} p_A(z_A, \rho(\tau))h_A(\rho(\tau))\rho'(\tau). \tag{4.57}$$

Differentiate each side of (4.50) with respect to τ, we obtain the identity

$$h_B(\tau) = \frac{h_B(L_B)}{h_A(L_A)} h_A(\rho(\tau))\rho'(\tau), \quad \forall \tau \in W_B. \tag{4.58}$$

Substituting (4.58) into (4.57) we have that

$$\int_{E_B}^{\tau} f_{A:B}^{*}(z_A)d\mu = \int_{E_B}^{\tau} z_B^{L} d\mu + p_A(z_A,\rho(\tau)) \int_{E_B}^{\tau} (z_B^{E} - z_B^{L})d\mu, \quad \forall \tau \in W_B$$

which is equation (4.42). This concludes the proof.■

Our model of V_{AB}^{*} is now complete. We define $f_{A:B}^{*}(z_A)$ implicitly by equation (4.42) with ρ chosen to satisfy implicitly the equation (4.50) and then set $V_{AB}^{*} = f_{A:B}^{*}(z_A)$. As a result of this choice for $f_{A:B}^{*}(z_A)$, we know by Proposition (4.3.1) that it also satisfies (4.49).

We make one final comment. Consider the restriction imposed on the choice for z_B given a particular choice for z_A. As we have stated before, this restriction is reflected in the inventory balance constraint

$$\int_{E_B}^{\tau} z_B d\mu \leq \int_{E_B}^{\tau} V_{AB}^{*} d\mu, \quad \forall \tau \in W_B. \tag{4.59}$$

Substituting in (4.59) the definition for V_{AB}^{*} we have that, $\forall \tau \in W_B$,

$$\int_{E_B}^{\tau} z_B d\mu \leq \int_{E_B}^{\tau} V_{AB}^{*} d\mu = \int_{E_B}^{\tau} f_{A:B}^{*}(z_A)d\mu = \int_{E_B}^{\tau} z_B^{L} d\mu + p_A(z_A,\rho(\tau)) \int_{E_B}^{\tau} (z_B^{E} - z_B^{L})d\mu$$

or, after rearranging terms,

$$p_B(z_B,\tau) \leq p_A(z_A,\rho(\tau)). \tag{4.60}$$

Essentially, Leachman and Boysen's method for modeling the dependence relationship between A and B for a Parallel $A:B$ Subnetwork was to require that the choices for z_A and z_B satisfy

(4.60).

4.3.4. Constructing an Independent Model of V_A for the Parallel $A{:}BC$ Subnetwork

For the Parallel $A{:}BC$ Subnetwork, the intermediate product transfer variable V_A was determined from the induced operating intensity via the intermediate product transfer functional

$$f_{A:BC}: Z_{a_1} \times \cdots \times Z_{a_M} \to d_1 Z_B + d_2 Z_C$$

(see 4.26 for the definition of $f_{A:BC}$). In this section, we will construct an independent model of V_A which we denote by V_A^*. The essential idea is to construct an $f_{A:BC}^*: D_A \to d_1 D_B + d_2 D_C$ which reasonably approximates $f_{A:BC}$ on $Z_{a_1} \times \cdots \times Z_{a_M}$ and then define $V_A^* = f_{A:BC}^*(z_A)$ for $z_A \in D_A$.

An inspection of the definition for the intermediate product transfer functional $f_{A:BC}$ shows that it satisfies the following boundary conditions, $\forall (z_{a_1}, \ldots, z_{a_M}) \in Z_{a_1} \times \cdots \times Z_{a_M}$, and $\min(E_B, E_C) \leq \tau \leq \max(L_B, L_C)$,

$$\int_{\min(E_B,E_C)}^{\max(L_B,L_C)} f_{A:BC}(z_{a_1}, \ldots, z_{a_M})\,d\mu = 1 \tag{4.61}$$

$$\int_{\min(E_B,E_C)}^{\tau} f_{A:BC}(z_{a_1}^L, \ldots, z_{a_M}^L)\,d\mu \leq \int_{\min(E_B,E_C)}^{\tau} f_{A:BC}(z_{a_1}, \ldots, z_{a_M})\,d\mu$$

$$\leq \int_{\min(E_B,E_C)}^{\tau} f_{A:BC}(z_{a_1}^E, \ldots, z_{a_M}^E)\,d\mu \tag{4.62}$$

For notational convenience, let $E_{BC} = \min(E_B, E_C)$ and $L_{BC} = \max(L_B, L_C)$ and $W_{BC} = [E_{BC}, L_{BC}]$. It is easily verified by the definition of $f_{A:BC}$ that

$$f_{A:BC}(z_{a_1}^L, \ldots, z_{a_M}^L) = d_1 z_B^L + d_2 z_C^L \tag{4.63}$$

$$f_{A:BC}(z^E_{a_1}, \ldots, z^E_a) = d_1 z^E_B + d_2 z^E_C$$

Of course, boundary conditions 4.61-4.62. are exactly analogous to the boundary conditions for $f_{A:B}$ given in 4.33-4.34. Based on the motivation and explanation provided for the development of the definition for $f^*_{A:B}$, we define $f^*_{A:BC}$ implicitly by the equation, $\forall \tau \in W_{BC}$,

$$\int_{E_{BC}}^{\tau} f^*_{A:BC}(z_A)\,d\mu = \int_{E_{BC}}^{\tau} f_{A:BC}(z^L_A)\,d\mu + p_A(z_A, \rho(\tau)) \int_{E_{BC}}^{\tau} [f_{A:BC}(z^E_A) - f_{A:BC}(z^L_A)]\,d\mu$$

$$= \int_{E_{BC}}^{\tau} [d_1 z^L_B + d_2 z^L_C]\,d\mu + p_A(z_A, \rho(\tau)) \int_{E_{BC}}^{\tau} [d_1(z^E_B - z^L_B) + d_2(z^E_C - z^L_C)]\,d\mu$$

for some $\rho : W_{BC} \to W_A$ which satisfies

$$\rho(E_{BC}) = E_A\,, \quad \rho(L_{BC}) = L_A$$

and which is

continuous and increasing.

Analogous to the additional requirement imposed on $f^*_{A:B}$ (see 4.49), let us insist that, $\forall$ $z_A \in D_A$, $f^*_{A:BC}$ also satisfies the equation, $\forall \tau \in W_{BC}$,

$$\frac{\displaystyle\int_{E_A}^{\rho(\tau)} \int_{E_A}^{x} (z_A - z^L_A)\,d\mu\,dx}{\displaystyle\int_{E_A}^{\rho(\tau)} \int_{E_A}^{x} (z^E_A - z^L_A)\,d\mu\,dx} = \frac{\displaystyle\int_{E_{BC}}^{\tau} \int_{E_{BC}}^{x} [f^*_{A:BC}(z_A) - (d_1 z^L_B + d_2 z^L_C)]\,d\mu\,dx}{\displaystyle\int_{E_{BC}}^{\tau} \int_{E_{BC}}^{x} [d_1(z^E_B - z^L_B) + d_2(z^E_C - z^L_C)]\,d\mu\,dx}.$$

By a proof similar to the proof of Proposition (4.3.1) it may be readily verified that ρ must satisfy the equation

$$\frac{\displaystyle\int_{E_{BC}}^{\tau} [d_1 h_B(x) + d_2 h_C(x)]\,dx}{d_1 h_B(L_B) + d_2 h_C(L_C)} = \frac{\displaystyle\int_{E_A}^{\rho(\tau)} h_A(x)\,dx}{h_A(L_A)}. \qquad (4.64)$$

To complete the model of V_A^* requires us to select the weights d_1 and d_2. To motivate our choice for d_1 and d_2, let us first consider the restriction imposed on the choices for z_B and z_C given a particular choice for z_A. As we have stated before, this restriction is reflected in the inventory balance constraint

$$\int_{E_{BC}}^{\tau} \{d_1 z_B + d_2 z_C\} d\mu \leqslant \int_{E_{BC}}^{\tau} V_A^* d\mu = \int_{E_{BC}}^{\tau} f_{A:BC}^*(z_A) d\mu , \quad \forall \tau \in W_{BC} . \tag{4.65}$$

Substituting into (4.65) the definition for $\int_{E_{BC}}^{\tau} f_{A:BC}^*(z_A) d\mu$ we obtain, after re-arranging terms, the inequality

$$\frac{\int_{E_{BC}}^{\tau} \{d_1(z_B - z_B^L) + d_2(z_C - z_C^L)\} d\mu}{\int_{E_{BC}}^{\tau} \{d_1(z_B^E - z_B^L) + d_2(z_C^E - z_C^L)\} d\mu} \leqslant p_A(z_A, \rho(\tau)) , \quad \forall \tau \in W_{BC} . \tag{4.66}$$

(4.66) is equivalent to the inequality

$$p_A(z_A, \rho(\tau)) \geqslant d_1 \left\{ \frac{h_B(\tau)}{d_1 h_B(\tau) + d_2 h_C(\tau)} \right\} p_B(z_B, \tau)$$

$$+ d_2 \left\{ \frac{h_C(\tau)}{d_1 h_B(\tau) + d_2 h_C(\tau)} \right\} p_C(z_C, \tau) , \quad \forall \tau \in W_{BC} . \tag{4.67}$$

Let A_1 denote the set of activities in A which precede the activities in B, let A_2 denote the set of activities in A which precede activities in C, and define

$$\alpha_{A_j A} = \frac{\sum_{l \in A_j} \alpha_{Al}}{\sum_{l \in A} \alpha_{Al}} , \quad j = 1,2 . \tag{4.68}$$

It may be easily derived algebraically that if $z_A = z_A^S$ for some feasible schedule S, then

$$p_A(z_A,\rho(\tau)) = \alpha_{A_1A}\left\{\frac{h_{A_1}(\rho(\tau))}{\alpha_{A_1A}(\rho(\tau))+\alpha_{A_1A}h_{A_2}(\rho(\tau))}\right\}p_{A_1}(z_{A_1}^S,\rho(\tau))$$

$$+\alpha_{A_2A}\left\{\frac{h_{A_2}(\rho(\tau))}{\alpha_{A_1A}h_{A_1}(\rho(\tau))+\alpha_{A_2A}h_{A_2}(\rho(\tau))}\right\}p_{A_2}(z_{A_2}^S,\rho(\tau)). \tag{4.69}$$

To motivate our choice for d_1 and d_2, consider the special case when, for some $l > 0$,

$$z_B^E(\tau) = z_{A_1}^E(\tau-l) = z_{A_2}^E(\tau-l) = z_C^E(\tau), \quad \forall\tau \in W_{BC} \tag{4.70}$$

$$z_B^L(\tau) = z_{A_1}^L(\tau-l) = z_{A_2}^L(\tau-l) = z_C^L(\tau), \quad \forall\tau \in W_{BC}.$$

In this case, an inspection of (4.64) shows that $\rho(\tau) = \tau - l$ (which is intuitively clear). Furthermore, $\rho(\tau) = \tau - l$ and (4.70) imply that

$$h_B(\tau) = h_{A_1}(\rho(\tau)) = h_{A_2}(\rho(\tau)) = h_C(\tau), \quad \forall\tau \in W_{BC}. \tag{4.71}$$

Substituting the identities found in (4.69) and (4.71) into the inequality (4.67) gives us the inequality, $\forall\tau \in W_{BC}$,

$$0 \geq \{d_1p_B(z_B,\tau)-\alpha_{A_1A}p_{A_1}(z_{A_1}^S,\rho(\tau))\} + \{d_2p_C(z_C,\tau)-\alpha_{A_2A}p_{A_2}(z_{A_2}^S,\rho(\tau))\}. \tag{4.72}$$

Since the subnetwork associated with A_1 and B (resp., A_2 and C) comprises a Parallel A_1:B Subnetwork (resp., a Parallel A_2:C Subnetwork), it follows from our work for that subnetwork that we would like to constrain z_B and z_C by the inequalities

$$p_B(z_B,\tau) \leq p_{A_1}(z_{A_1}^S,\rho(\tau)), \quad \tau \in W_{BC} \tag{4.73}$$

$$p_C(z_C,\tau) \leq p_{A_2}(z_{A_2}^S,\rho(\tau)), \quad \tau \in W_{BC}.$$

Hence, if we set

$$d_j = \alpha_{A_jA}, \quad j = 1,2 \tag{4.74}$$

then any $z_B \in D_B$, $z_C \in D_C$ which *do* satisfy (4.73) will satisfy (4.72). Hence, for this choice of d_1 and d_2 (and for this special case), we will not eliminate any reasonable choices for z_B and z_C. Therefore, we specify d_1 and d_2 by (4.74).

We make one final remark about the model for V_A^* and hence our model for the dependence relationship between A, B, and C. Leachman and Boysen's method for modeling the dependence relationship between A, B, and C is to insist that z_A, z_B, and z_C satisfy the inequality

$$\alpha_{A_1A} p_B(z_B, \rho_B^{-1}(\tau)) + \alpha_{A_2A} p_C(z_C, \rho_C^{-1}(\tau)) \leq p_A(z_A, \tau), \quad \forall \tau \in W_{BC} \tag{4.75}$$

where, for $i = B, C$, $\rho_i^{-1}(\tau)$ satisfies

$$\frac{\int_{E_A}^{\tau} h_A(x)\,dx}{h_A(L_A)} = \frac{\int_{E_i}^{\rho_i^{-1}(\tau)} h_i(x)\,dx}{h_i(L_i)}.$$

An inspection of the relationship between (4.75) and (4.72). reveals that, except for the differences in the definition of $\rho(\tau)$, *Leachman and Boysen's method for modeling the dependence relationship is our method under the special case cited above.* For more general cases, we feel our approach is more sensible since it will weight α_{AA_1} and α_{AA_2} by *time-varying* factors. Ultimately, it will have to be tested to see if it performs better. (See Section (4.5.4) for a discussion concerning testing of the approach for modeling the flow types.)

4.3.5. Constructing an Independent Model of the Production Function

For both subnetworks analyzed, $F_A(y_A, W_A)$ does satisfy the usual boundary conditions. It would therefore be appropriate to model it in exactly the same manner as we modeled V_{AB}^* or V_A^*. Since V_{AB}^* and V_A^* were modeled in detail, we omit the analysis.

4.4. Analyzing Subnetworks Through the Technique of Replication

In Section 4.3, we developed models of the flow types for the Parallel $A:B$ Subnetwork and the Parallel $A:BC$ Subnetwork. This section introduces the technique of *replication* of detailed activities and uses it to develop models of the flow types for two further classes of subnetworks. We proceed to develop models of the flow types for two classes of *Non-Parallel* $A:BC$ Subnetworks.

4.4.1. The Complete Precedence Non-Parallel *ABC* Subnetwork

The Complete Precedence Non-Parallel ABC Subnetwork is shown in Figure (4-5). Here, detailed activities $a_1,a_2,\ldots,a_M$ were aggregated into aggregate A, detailed activities $b_1,b_2,\ldots,b_M$ were aggregated into aggregate B, and detailed activities $c_1,c_2,\ldots,c_M$ were aggregated into aggregate C. Let G denote the underlying production system and let G^* denote the production network associated with G. We will construct a production system H which is, in effect, "equivalent" to G and analyze H to arrive at models of the flow types for the subnetwork shown in Figure (4-5).

To construct H, let us first construct the production network H^* associated with H:

Step 1: Add nodes $a_{12},\ldots,a_{M2}$ to G. Re-label node a_i to a_{i1}, $i=1,2,\ldots,M$.

Step 2: Add arc (d,a_{i2}) to G^* if there is an arc (d,a_i) in G, $i=1,2,\ldots,M$.
Add arc (a_{i2},d) to G^* if there is an arc (a_i,d) in G, $i=1,2,\ldots,M$.

H^* is the network obtained from G^* from Steps 1 and 2. The subnetwork in H which corresponds to the Complete Precedence Non-Parallel $A:BC$ Subnetwork in G is shown in Figure (4-6). As notation, we have let the symbol A_j, $j=1,2$, denote the aggregate activity in H comprised of activities a_{ij}, $i=1,2,\ldots,M$. Essentially, H^* is G^* with nodes $a_1,\ldots,a_M$ "replicated."

The production system H whose production network is H^* is defined as follows:

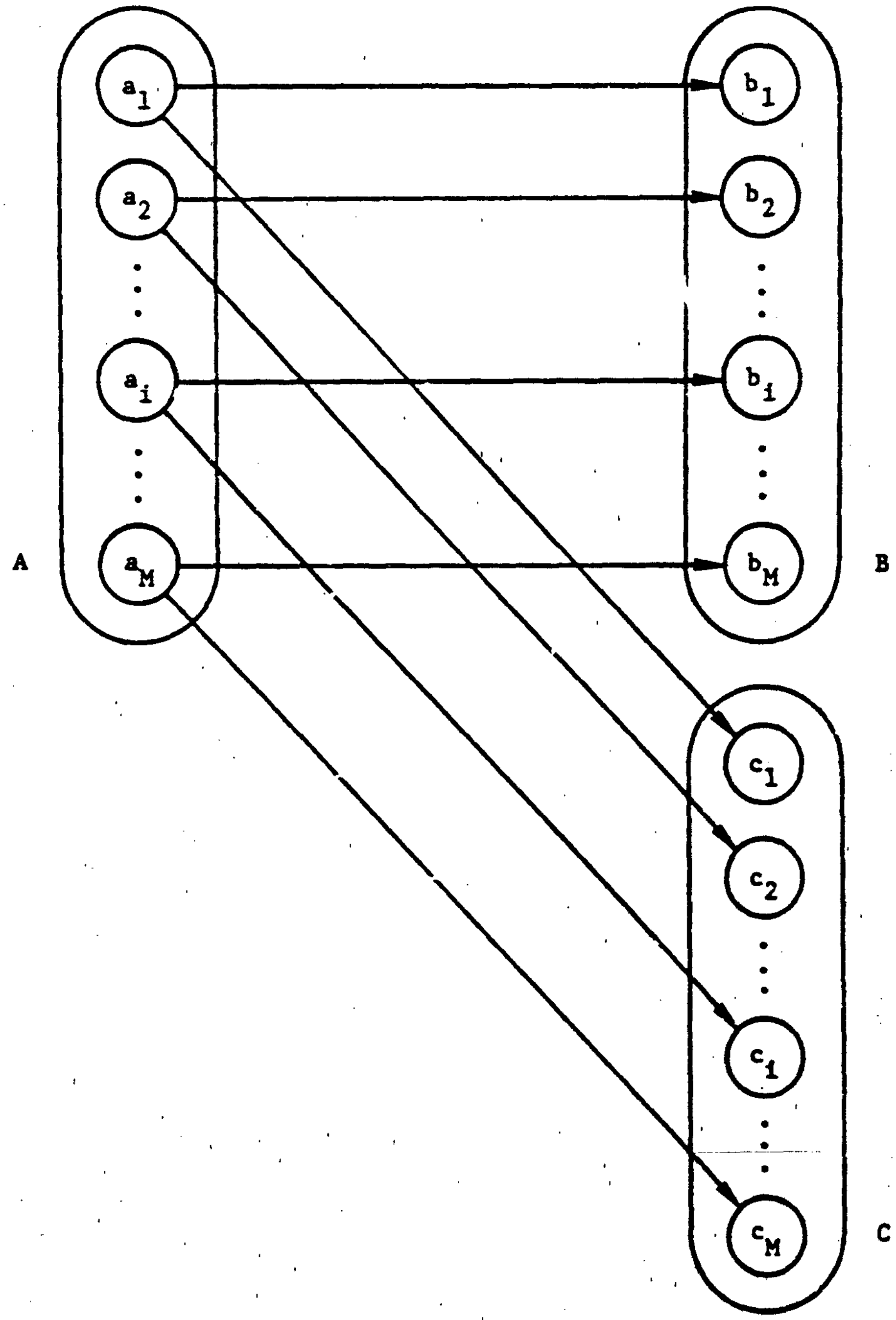


FIGURE (4-5)

THE COMPLETE PRECEDENCE, NON-PARALLEL A:BC SUBNETWORK

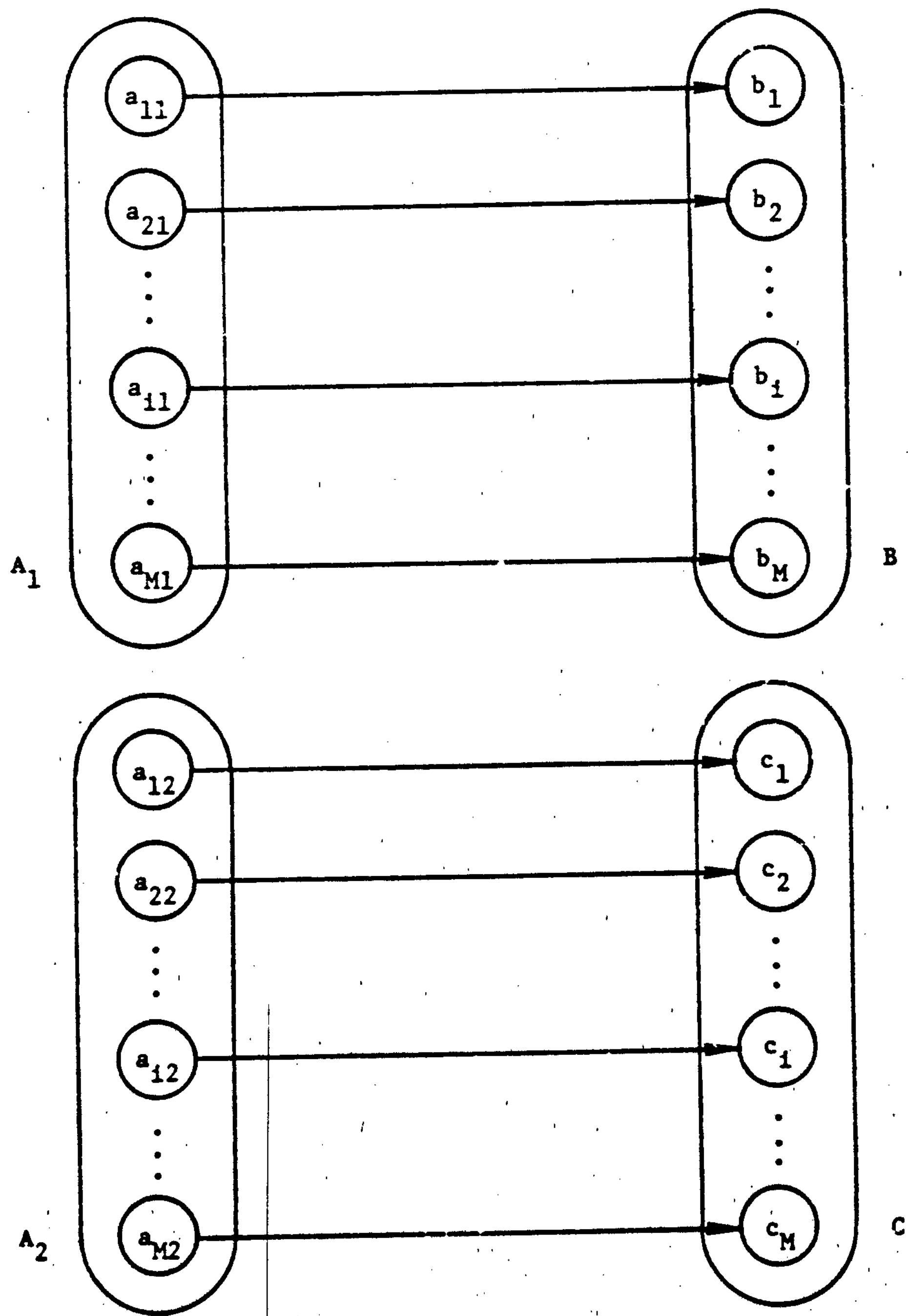


FIGURE (4-6)

SUBNETWORK IN H CORRESPONDING TO

COMPLETE PRECEDENCE A:BC SUBNETWORK IN G

(i) The duration and resource requirements of activity a_{ij} in H are identical to the duration and resource requirements of activity a_i in G, $i = 1,2,\ldots,M$, $j = 1,2$.

(ii) For an activity in H other than some a_{ij}, the duration and resource requirements are identical to the duration and resource requirements of its corresponding counterpart in G.

(iii) It is required that the operating intensities $z_{a_{i1}}$ and $z_{a_{i2}}$ are equal, $i = 1,2,\ldots,M$.

Since the subnetwork pictured in Figure (4-6) is comprised of two Parallel $A{:}B$ Subnetworks, it follows that for the production system H, the variables $F_{A_1}(z_{A_1})$, $F_{A_2}(z_{A_2})$, V_{A_1B}, V_{A_2C}, W_B, and W_C are determined from the induced operating intensities z_{A_1} and z_{A_2} exactly as described in Section 4.2.1. Given $z_{A_j} \in Z_{A_j}$, $j = 1,2$, the restrictions imposed on the choices for z_B and z_C are modeled by the constraints

$$\int_{E_B}^{\tau} (V_{A_1B} - z_B)\,d\mu \geq 0, \quad \tau \in W_B \tag{4.76}$$

$$\int_{E_C}^{\tau} (V_{A_2C} - z_C)\,d\mu \geq 0, \quad \tau \in W_C \tag{4.77}$$

It is immediate by our construction of H and (ii) above in particular that Z_A, the set of induced operating intensities for aggregate A in G, is equal to Z_{A_j}, $j = 1,2$, the set of induced operating intensities for aggregate A_j in H, $j = 1,2$, when these sets are viewed as sets of functions. If one identifies Z_A with Z_{A_j}, $j = 1,2$, then it is immediate by our method for modeling the intermediate product transfers presented in Section 4.2 that

$$V_{AB} = V_{A_1B}, \quad V_{AC} = V_{A_2C}. \tag{4.78}$$

Hence, it is appropriate to view aggregate A in G as producing *two* identical products, i.e., to

define for $z_A \in Z_A$

$$F_A(y_A, W_A) = \left[F_A^B(y_A, W_A),\ F_A^C(y_A, W_A)\right] = \left[F_{A_1}(y_A, W_A),\ F_{A_2}(y_A, W_A)\right] \tag{4.79}$$

(where F_{A_i} is defined by (4.9)) and, in view of (4.79), to define for $z_A \in Z_A$

$$V_{AB} = f_{A_1 B}(z_{a_1}, \ldots, z_{a_M}) = V_{A_1 B}\,,\quad V_{AC} = f_{A_2 C}(z_{a_1}, \ldots, z_{a_M}) = V_{A_2 C}\,. \tag{4.80}$$

The restrictions on the choices for z_B and z_C given z_A would be modeled by constraints 4.77-4.78 (with the substitution of the identities given in 4.81).

Since the models of the flow types determined from the induced operating intensities for these two subnetworks are identical, it seems reasonable to insist that the independent models of the flow types determined from the aggregate intensities also be identical. Thus, we define V_{AB}^*, V_{AC}^*, and $F_A^*(y_A, W_A)$ to be

$$V_{AB}^* = V_{A_1 B}^*$$

$$V_{AC}^* = V_{A_2 C}^*$$

$$F_A^*(y_A, W_A) = \left[F_{A_1}^*(y_A, W_A),\ F_{A_2}^*(y_A, W_A)\right].$$

The variables $F_{A_i}^*(y_A, W_A)$, $i = 1,2$, $V_{A_1 B}^*$ and $V_{A_2 C}^*$ were defined in Section 4.3.2. This concludes our analysis of this subnetwork.

4.4.2. The Partial Precedence Non-Parallel $A{:}BC$ Subnetwork

The *Partial Precedence Non-Parallel $A{:}BC$ Subnetwork* is shown in Figure (4-7). Here, detailed activities $a_1, \ldots, a_M$ were aggregated into aggregate A, detailed activities $b_1, \ldots, b_M$ were aggregated into aggregate B, and detailed activities $c_1, \ldots, c_L$ were aggregated into aggregate C. Let G denote the underlying production system and let G^* denote the production network associated with G. We will construct a production system H which is, in effect, equivalent

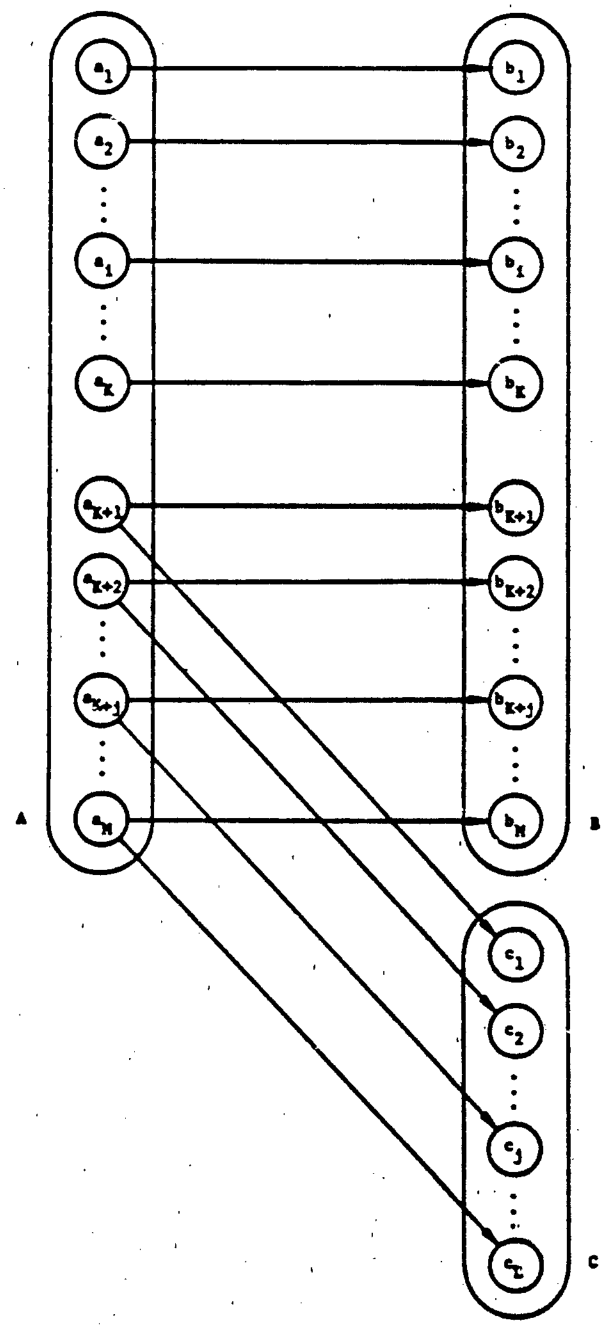


FIGURE (4-7)

THE PARTIAL PRECEDENCE NON-PARALLEL A:BC SUBNETWORK

to G and analyze H to arrive at models of the flow types for the subnetwork shown in Figure (4-7).

To construct H, let us first construct the production network H^* associated with H:

Step 1: Add nodes $a_{(k+1)2}, \ldots, a_{M2}$ to G^*. Re-label node a_i to a_{i1}, $i = K+1, \ldots, M$.

Step 2: Add arc (d, a_{i2}) to G^* if there is an arc (d, a_i) in G, $i = K+1, \ldots, M$. Add arc (a_{i2}, d) to G^* if there is an arc (a_i, d) in G, $i = K+1, \ldots, M$.

H^* is the network obtained from G^* from Steps 1 and 2. The subnetwork in H which corresponds to the Partial Precedence Non-Parallel $A{:}BC$ Subnetwork in G is shown in Figure (4-8). As notation, we have let the symbol A_j, $j = 1,2$, denote the aggregate activity in H comprised of activities a_{ij}, $i = K+1, \ldots, M$. Essentially, H^* is G^* with nodes $a_{K+1}, \ldots, a_M$ "replicated."

The production system H whose production network is H^* is defined as follows:

(i) The duration and resource requirements of activity a_{ij} in H are identical to the duration and resource requirements of activity a_i in G, $i = K+1, \ldots, M$, $j = 1,2$.

(ii) For an activity in H other than some a_{ij}, the duration and resource requirements are identical to the duration and resource requirements of its corresponding counterpart in G.

(iii) It is required that the operating intensities $z_{a_{i1}}$, $z_{a_{i2}}$ are equal, $i = K+1, \ldots, M$.

It is immediate by our construction of H that $Z_{A'}$, the set of induced operating intensities for aggregate A' in H, equals Z_A, the set of induced operating intensities for aggregate A in G, and that Z_{A_2}, the set of induced operating intensities for aggregate A_2 in H, equals Z_{A_1}, the set of induced operating intensities for A_1 in G (when all sets are viewed as sets of functions). If one identifies $Z_{A'}$ with Z_A and Z_{A_2} with Z_{A_1} then it is immediate by our modeling of the intermediate product transfers presented in Section 4.2 that, for $z_A \in Z_A$,

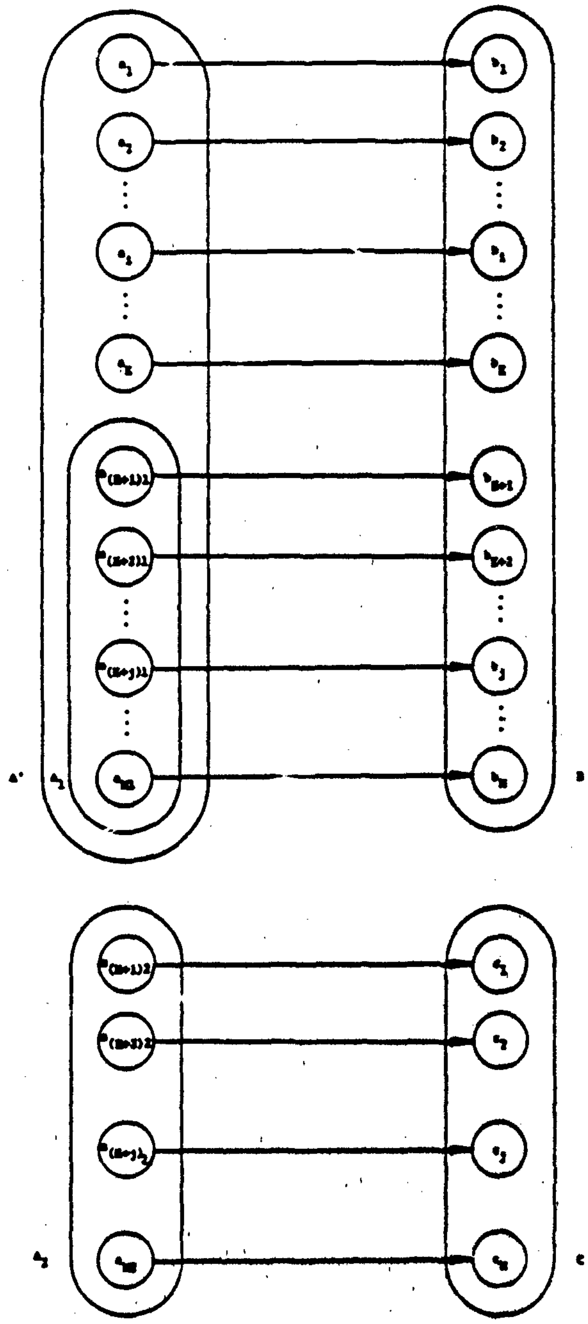


FIGURE (4-8)

SUBNETWORK IN H CORRESPONDING TO

PARTIAL PRECEDENCE A:BC SUBNETWORK IN G

$$V_{AB} = f_{A':B}(z_{a_1}, \ldots, z_{a_{(K+1)1}}, \ldots, z_{a_{M1}})$$

$$V_{AC} = V_{A_1C} = f_{A_2:C}(z_{a_{(K+1)2}}, \ldots, z_{a_{M2}})$$

$$F_A(y_A, W_A) = \sum_{i=1}^{i=K} \alpha_{Bb_i} z_{a_i} + \sum_{i=K+1}^{M} \alpha_{Cc_i} z_{c_i}.$$

Since the models of the intermediate product transfer flow types determined from the induced operating intensities for these two subnetworks are identical, it seems reasonable to insist that the independent models of these flow types determined from the aggregate operating intensity also be identical. When A_2 and A' have been aggregated, the subnetwork pictured in Figure (4-8) is a Parallel $A:BC$ Subnetwork. From having studied this case in Section 4.3, we know that while it is not possible to arrive at an independent model for either $V_{A'B}$ or V_{A_2C} we do know how to constrain the choices for z_B and z_C given z_{A^*} by modeling the sum $V_{A^*} = V_{A'B} + V_{A_2C}$. That is, z_B and z_C must satisfy (see 4.67)

$$p_{A^*}(z_{A^*}, \rho(\tau)) \geq d_1 \left\{ \frac{h_B(\tau)}{d_1 h_B(\tau) + d_2 h_C(\tau)} \right\} p_B(z_B, \tau)$$

$$+ d_2 \left\{ \frac{h_C(\tau)}{d_1 h_B(\tau) + d_2 h_C(\tau)} \right\} p_C(z_C, \tau), \quad \forall \tau \in W_{BC} \tag{4.81}$$

where

$$d_1 = \alpha_{A'A^*} = \frac{\sum_{i \in A'} \alpha_{A^*i}}{\sum_{i \in A^*} \alpha_{A^*i}}$$

$$d_2 = \alpha_{A_2A^*} = \frac{\sum_{i \in A_2} \alpha_{A_2i}}{\sum_{i \in A^*} \alpha_{A^*i}}$$

$$z_{A^*}^E = \alpha_{A'A^*} z_{A'}^E + \alpha_{A_2A^*} z_{A_2}^E$$

$$= \alpha_{A'A^*} z_A^E + \alpha_{A_2A^*} z_{A_1}^E$$

$$z^L_{A''} = \alpha_{A'A''} \cdot z^L_{A'} + \alpha_{A_2A''} \cdot z^L_{A_2}$$

$$= \alpha_{A'A''} \cdot z^L_{A} + \alpha_{A_2A''} \cdot z^L_{A_1}$$

and $\rho(\tau)$ satisfies the equation, $\forall \tau \in W_{BC}$,

$$\frac{\int_{\Sigma_{BC}}^{\tau} \{d_1 h_B(x) + d_2 h_C(x)\} dx}{d_1 h_B(L_B) + d_2 h_C(L_C)} = \frac{\int_{\Sigma_{A}}^{\rho(\tau)} \{d_1 h_{A'}(x) + d_2 h_{A_2}(x)\} dx}{d_1 h_{A'}(L_{A'}) + d_2 h_{A_2}(L_{A_2})}$$

$$= \frac{\int_{\Sigma_{A}}^{\rho(\tau)} \{d_1 h_{A}(x) + d_2 h_{A_1}(x)\} dx}{d_1 h_{A}(L_{A}) + d_2 h_{A_2}(L_{A_2})}.$$

Note that if $z_{A''} = z^S_{A''}$ for some feasible schedule S in H then by (4.69) we would have that, $\forall$ $\tau \in W_{BC}$,

$$p_{A''}(z_{A''}, \rho(\tau)) = d_1 \left\{ \frac{h_{A'}(\rho(\tau))}{d_1 h_{A'}(\rho(\tau)) + d_2 h_{A_2}(\rho(\tau))} \right\} p_{A'}(z_{A'}, \rho(\tau))$$

$$+ d_2 \left\{ \frac{h_{A_2}(\rho(\tau))}{d_1 h_{A'}(\rho(\tau)) + d_2 h_{A_2}(\rho(\tau))} \right\} p_{A_2}(z_{A_2}, \rho(\tau))$$

$$= d_1 \left\{ \frac{h_{A}(\rho(\tau))}{d_1 h_{A}(\rho(\tau)) + d_2 h_{A_1}(\rho(\tau))} \right\} p_{A}(z_{A}, \rho(\tau))$$

$$+ d_2 \left\{ \frac{h_{A_1}(\rho(\tau))}{d_1 h_{A}(\rho(\tau)) + d_2 h_{A_1}(\rho(\tau))} \right\} p_{A_1}(z_{A_1}, \rho(\tau)).$$

Intuitively, p_B and p_C are being constrained by p_A and p_{A_1} which makes sense based on our analysis of Parallel $A:B$ Subnetworks.

If one used $z_{A''}$, then (4.81) would serve to model the dependence relationships. Unfortunately, it is $y_{A'}$ not $y_{A''}$ that models resource use for A. Without knowing z_{A_1} we could not easily transform from $y_{A''}$ to y_A. However, if we made the simplifying assumption that

$$z_{A''} = z_{A'} = z_{A_2} \; (= z_A = z_{A_1}) \tag{4.82}$$

then

$$y_{A''} = y_A \text{ and } p_{A''} = p_A \, .$$

Essentially, Leachman and Boysen's method for modeling the dependence relationships for this subnetwork is to use (4.82) with the simplifying assumption (4.83). While (4.83) is intuitively appealing-- without additional information, it assumes all activities within aggregate A progress at the same rate-- it is restrictive. More than anything else, it points out the fact that modeling dependence relationships for complicated (i.e., non-parallel) subnetworks is difficult.

4.5. Concluding Remarks

4.5.1. Extensions

The development of the models of the flow types for the Parallel $A:B$, Parallel $A:BC$, Complete Precedence Non-Parallel $A:BC$, and the Partial Precedence $A:BC$ Subnetworks presented in Sections 4.1-4.4 may be repeated for similar classes of subnetworks. For example, it is easy to see by "symmetry" how one would model the flow types for the Parallel $BC:A$ Subnetwork (see Figure (4-9)), the Complete Precedence Non-Parallel $BC:A$ Subnetwork (see Figure (4-10)), and the Partial Precedence Non-Parallel $BC:A$ Subnetwork. The subnetworks shown in Figures (4-9), (4-10) and those analyzed in detail in this chapter include all of the subnetworks presented in the Leachman and Boysen [1983] paper.

4.5.2. Constructing Tractable Models

In order to develop a convenient method for accomplishing multi-project resource-use planning, it is desirable to construct models of the flow types which are *tractable*. That is, the constraints which define the set Z of feasible choices for the allocations of resources to the aggregate activities induced from the models of the flow types are linear. An inspection of the models of the flow types developed in Section 4.2-4.4 reveals that only the indefinite integrals of the flow types and not the flow types themselves, are required to determine the constraints which define the set Z. To define models of the flow types which are tractable, we simply construct a *piecewise-linear approximation* to the indefinite integrals of the flow types.

Definition (4.83)

Let $z \in L_+^\infty(\mu)$. The piecewise-linear approximation to $\int_0^x z d\mu$, denoted by $\int_0^x z^* d\mu$, is defined by

$$\int_0^x z^* d\mu = \int_0^{\lfloor x \rfloor} z d\mu + (x - \lfloor x \rfloor) \left\{ \int_0^{\lfloor x \rfloor + 1} z d\mu - \int_0^{\lfloor x \rfloor} z d\mu \right\}$$

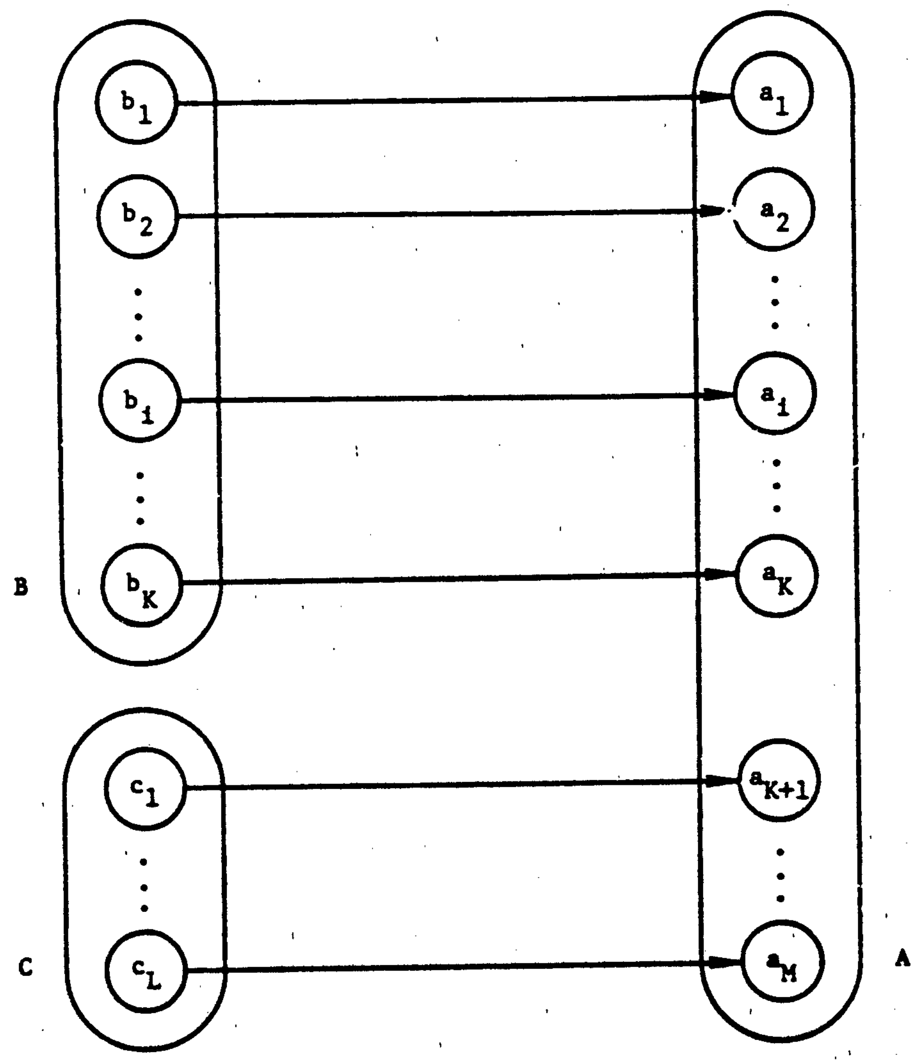


FIGURE (4-9)

THE PARALLEL BC:A SUBNETWORK

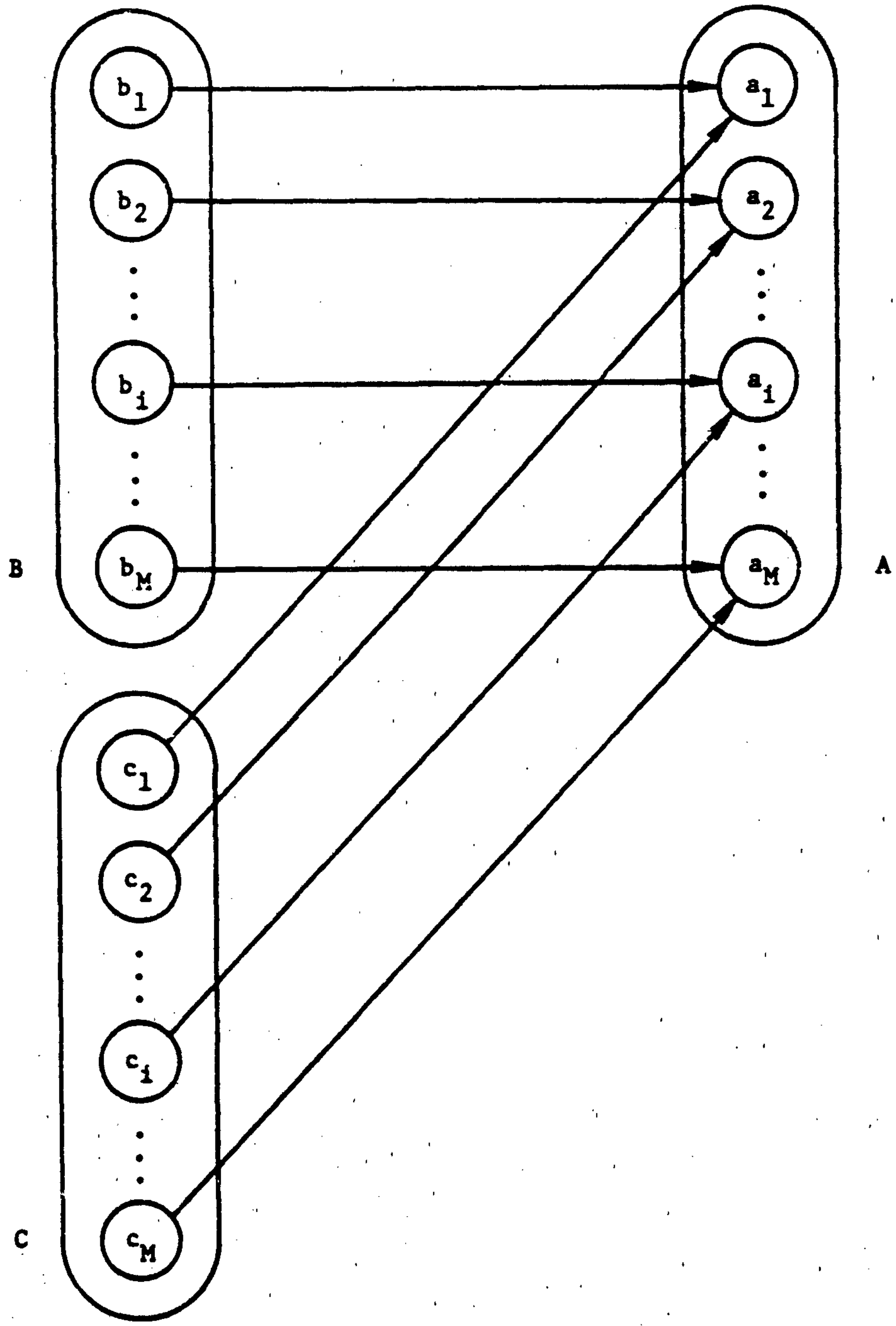


FIGURE (4-10)

THE COMPLETE PRECEDENCE, NON-PARALLEL BC:A SUBNETWORK

(where $\lfloor x \rfloor$ equals the greatest integer of x).

The models of the flow types are now understood to be the piecewise-linear approximations to the previous models of the flow types. (Since the aggregate operating intensities were constrained to be step-functions, the indefinite integrals of such flow types are already piecewise-linear.)

An inspection of the constraints which model the dependence relationships between aggregate activities given in Section 4.3 clearly shows that they now become linear constraints. Since all flow types and hence all constraints which model dependence relationships are determined from the aggregate operating intensities, a Linear Program which accomplishes multi-project resource-use planning may be formulated by having the indefinite integrals of the aggregate operating intensities as the decision variables. A suitable objective function (one for which minimizes cost, for example) is all that is required. See Leachman and Boysen [1983] for their choice for an objective function.

4.5.3. Lack of Consistency for the Models of the Flow Types

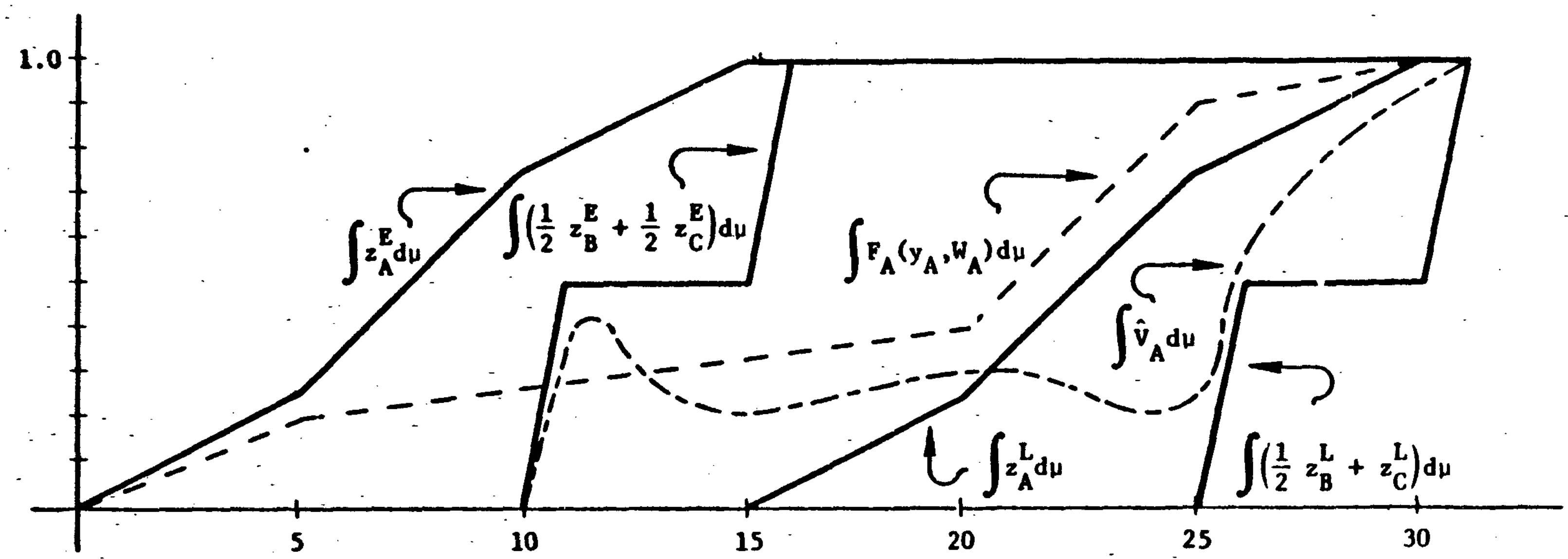
The constraints which model the dependence relationships are required to be satisfied and appear in the Linear Program discussed in Section (4.5.1). However, it is possible that the inventory balance constraints which link the production variable to the intermediate product transfer variables are not satisfied.

Consider the example of a Parallel $A:BC$ Subnetwork (where A has two activities) shown in Figure (4-11). The inventory balance constraint

$$\int_{E_A}^{\tau} [F_A^*(y_A, W_A) - V_A^*] d\mu \geq 0 \qquad (4.84)$$

is not always satisfied. Furthermore, the indefinite integral of V_A^*,

$$\int_{E_A}^{\tau} V_A^* d\mu \qquad (4.85)$$



	Early-starts	Late-starts	Durations	Resource Requirements
a_1	0	15	10	$\alpha_{Aa_1} = .5$
a_2	5	20	10	$\alpha_{Aa_2} = .5$
B	10	25	1	-
C	15	30	1	-

FIGURE (4-11)

EXAMPLE OF INCONSISTENCY OF MODELS

is not always increasing, i.e., $\dot{V}_A$ is negative at some points in time.

The main reasons why such an example can be constructed are twofold. First, the domain D_A is too large, i.e., additional constraints must be imposed to insure that the variable z_A is close to an induced operating intensity (the particular z_A in Figure (4-11) is not).[1] Second, the boundaries z_B^E and z_B^L are completely different in "shape" than the boundaries z_A^E and z_A^L. Perhaps such inconsistencies point to the fact that the underlying activities should not be aggregated.

In any event, models of the flow types associated with the production functions and the intermediate product transfers should satisfy the inventory balance constraints similar to (4.85). In addition, an inconsistency such as (4.86) should not occur either. As a test for the construction of the models, these criteria could be employed.

4.5.4. Suggestions for Future Research

It is desirable to extend the analyses provided in Sections 4.2-4.4 to wider classes of more complicated networks. In addition, the approach to constructing independent models (Section 4.3) which is the key to modeling the dependence relationships should be tested. Results of a preliminary test of the approach may be found in Dalebout's [1983] Master's thesis. Based on her results, the approach due to Leachman and Boysen, extended in this chapter, seems to work well. Lastly, further study into methods for restricting the domains of the aggregate operating intensities so as to insure consistency is desirable and might be the key to more realistic models of the dependence relationships.

[1] See Leachman and Boysen [1983] for their proposals for furter restrictions on the domain D_A.

REFERENCES

[1979] Bazarra, M. and C. M. Shetty, NONLINEAR PROGRAMMING, John Wiley & Sons, New York.

[1963] Berge, C., TOPOLOGICAL SPACES, Macmillan Co., New York.

[1959] Bunge, M., METASCIENTIFIC QUERIES, Charles C. Thomas Publisher, Springfield, Illinois.

[1973] Bunge, M., METHOD, MODEL AND MATTER, D. Reidel Publishing Co., Dordrecht, Holland.

[1983] Dalebout, D., "A Study of Dependence Relationships Between Aggregate Activities in a Project-Oriented Production System," Master's Thesis, Department of Industrial Engineering and Operations Research, University of California, Berkeley.

[1966] Dugundgi, J., TOPOLOGY, Allyn and Bacon, Boston, Massachusetts.

[1983] Hackman, S. T. and R. W. Shephard, "Production Networks: A Dynamic Model of Production," ORC Report 83-4, Operations Research Center, University of California, Berkeley.

[1974] Halmos, P. R., MEASURE THEORY, Springer-Verlag, New York.

[1972] Hanock, G. and M. Rothschild, "Testing the Assumptions of Production Theory: A Nonparametric Approach," *Journal of Political Economy*, **80.**

[1983] Kasper, M., "Production Planning of Capacitated, Multi-Stage Discrete Manufacturing Systems," Ph.D. Thesis, Department of Industrial Engineering and Operations Research, University of California, Berkeley.

[1955] Kelley, J. L., GENERAL TOPOLOGY, Springer-Verlag, New York.

[1951] Koopmans, T. C., ed., ACTIVITY ANALYSIS OF PRODUCTION AND ALLOCATION, John Wiley & Sons, New York.

[1982] Kurtulus, I. and E. W. Davis, "Multi-Project Scheduling: Categorization of Heuristic Rules Performance," *Management Science*, **28**, (2), 161-172.

[1982] Leachman, R. C., "Production Planning for Multi-Resource Network Systems," *Naval Research Logistics Quarterly*, **29**, (1), 47-54.

[1983] Leachman, R. C., "Multiple Resource Leveling in Construction Systems Through Variation of Activity Intensities," *Naval Research Logistics Quarterly*, **30**, (2), 187-198.

[1984] Leachman, R. C., "Dynamic Activity Analysis Model: Framework for Production Modeling," to be submitted to *Management Science.*

[1982] Leachman, R. C. and J. Boysen, "Aggregate Network Model of Ship Overhaul," ORC Report 82-2, Operations Research Center, University of California, Berkeley.

[1983] Leachman, R. C. and J. Boysen, "An Aggregate Model for Multi-Project Resource Allocation," PROJECT MANAGEMENT: METHODS AND STUDIES, North Holland, New York.

[1951] Leontief, W. W., THE STRUCTURE OF AN AMERICAN ECONOMY, Oxford University Press.

[1981] Mak, K. T., "Stochastic Theory of Production Correspondences," ORC Report 81-6, Operations Research Center, University of California, Berkeley.

[1970] Moder, J. J. and C. R. Phillips, PROJECT MANAGEMENT WITH CPM AND PERT, 2nd ed., Van Nostrand Reinhold, New York.

[1954] Morgenstern, O., ed., ECONOMIC ACTIVITY ANALYSIS, John Wiley & Sons, New York.

[1970] Orlicky, J. A., "Requirements Planning Systems: Cinderalla's Bright Prospects for the Future," Proceedings of the American Production and Inventory Control Society, 228-229.

[1975] Orlicky, J. A., MATERIAL REQUIREMENTS PLANNING, McGraw Hill, New York.

[1967] Plossl, G. W. and O. W. Wright, PRODUCTION AND INVENTORY CONTROL PRINCIPLES AND TECHNIQUES, Prentice Hall, Englewood Cliffs.

[1968] Royden, H. L., REAL ANALYSIS, MacMiillan Co., New York.

[1973] Rudin, W., FUNCTIONAL ANALYSIS, McGraw Hill, New York.

[1970a] Shephard R. W., THEORY OF COST AND PRODUCTION FUNCTIONS, Princeton University Press, Princeton, New York.

[1970b] Shephard, R. W., "Proof of the Law of Diminishing Return," Zeitschrift für Nationalökonomie, 30, 7-34.

[1977] Shephard, R. W., R. Al-Ayat and R. C. Leachman, "Shipbuilding Production Function: An Example of a Dynamic Production Function," in Quantitative Wirtschaftsforschung Festschrift Volume [Wilhelm Krelle] J.B..C. Mohr [ed.], Tübingen.

[1980] Shephard, R. W. and R. Färe, DYNAMIC THEORY OF PRODUCTION CORRESPONDENCES: Vol. 54, Mathematical Systems in Economics, Verlag Anton Hain, Meisenheim.

END

FILMED

10-85

DTIC